

The British Journal for the Philosophy of Science

VOLUME VII

FEBRUARY, 1957

No. 28

ON THE DIFFERENCE BETWEEN MEN AND MACHINES *

T. R. MILES

If you prick us, do we not bleed? If you tickle us, do we not laugh?—
The Merchant of Venice.

1 The Problem. 2 The 'Traditional' Answer. 3 The 'Verificationist' Answer. 4 The Concept of 'Body-schema'; a Suggested Synthesis. 5 Conclusions.

I *The Problem*

IN the normal way the question, 'What is the difference between a man and a machine?' presents no difficulty. There is no shortage of criteria for distinguishing them. We can tell at a glance that even the most ingenious electronic device is nothing like a man. It just does not look like one. It has no eyes or nose, and its surface is not made of skin. If we prick it, it does not bleed, and, if we tickle it, it does not laugh.

Even if a machine were sent on completion to Madame Tussaud and made to look like a human being, and even if people were at times deceived into thinking it a human being, as they do the policeman on the stairs at Madame Tussaud's, there is no doubt how the mistake could be corrected. Further examination would show that there were valves, or cog-wheels inside, and that wax had been used in place of skin. Moreover, as Kapp has pointed out very forcibly, machines have a so-called 'input function'. 'They are manipulated at more or less frequent intervals by an operator, who uses handles, levers, pedals, push-buttons, and similar controlling devices; and their performance is not a random one but one that the operator contemplates while he is exercising the control.'¹ In contrast human beings do not cease to function if other human beings fail to operate the requisite controls. But even if an object were constructed which did not require interference from an operator, and even if a suitable internal mechanism

* Received 27. vii.55

¹ R. O. Kapp, 'Living and lifeless machines', this *Journal*, 1954, 5, 98

made spontaneous movement possible, the visual appearance would almost certainly be of very 'wooden' movement. As Michotte¹ has shown, there is the appearance of 'live' movement only if the visual field is organised in certain special ways. Unless the requisite conditions were present in this particular case, there would be little doubt that we were looking at a machine and not a man.

If there is to be the sort of problem with which this paper is concerned, we must suppose the possibility of a flesh and blood machine capable of producing exactly the same responses as would be expected of a human being. Such an object would be constructed so that it could be graded A, B, etc. by the army medical authorities, so that its I.Q. could be determined by psychologists and its psychopathology by psychiatrists. It would show affection if we were kind to it; it would show anger and dismay if we tried to cheat it; and, if we confronted it with a beautiful poem or a beautiful sunset, it would make appropriate remarks of commendation. In brief, examination would yield all the results that would be expected of a human being within the normal limits of human variability. In what follows I shall refer to this object as a *homo mechanisma*.

I am not concerned in this paper to discuss whether a *homo mechanisma* is ever likely to be made, nor to raise any technical problems on how to make one. My purpose is rather to assume that one is possible in order to see how the concepts 'man', 'machine', 'mind' and 'consciousness' would operate if it were made. This procedure will, I hope, throw light on the way in which these concepts function at present, even though, here and now, *homo mechanisma* is a subject of fantastic speculation rather than of practical possibility.

For present purposes, therefore, it is enough that a *homo mechanisma* is possible *logically*. In other words, there is no self-contradiction in postulating an object of this sort. To say 'Please make me a *homo mechanisma*' would not be absurd in the way in which 'Please make me a round square' would be absurd. The supposition is in effect that there are no limits to what modern engineering may achieve, and that, whatever characteristically 'human' responses anyone chooses to name, these responses could be reproduced artificially by a sufficiently competent engineer.²

¹ A. Michotte, *La perception de la causalité*, Louvain, 1954, Chapter XII

² This supposition is, of course, a truism, rather than an expression of faith in scientific progress. Compare D. M. Mackay, 'Mentality in Machines', *Aristotelian Society Proceedings*, Suppl. Vol. 26, 1955, 85.

ON DIFFERENCE BETWEEN MEN AND MACHINES

Two further complications need to be mentioned. In the first place, if we were not sure whether an object in front of us was a man or a machine, it would be possible, at least in principle, to study its past history. We could thus tell whether it had come into existence artificially or by the normal processes of human reproduction. Anything in the latter group could then be labelled 'man', and anything in the former group could be labelled 'machine'. For purposes of argument it will be necessary to ignore distinguishing criteria that arise as a result of the object's past history.

Secondly, there is the problem of growth. Human beings grow, and though it is logically possible that a machine should show exactly similar signs of growth, this machine could still be distinguished from human beings if its growth was not of the normal 'from-cradle-to-grave' variety. To make the formulation of our problem more water-tight, let us assume that a *homo mechanisma* becomes 'married' to a human being, and that a 'child' is produced in the normal way, which grows up like any ordinary child. We may refer to this 'child' as a *second-generation homo mechanisma*.

The problem can now be stated as follows: *On what grounds, other than those of ancestry, could one assert of a second-generation homo mechanisma either that it was a man or that it was a machine?*¹

If a positive answer can be given to this question—and I shall claim that it can—it follows that there will always be a difference between man and machines, however similar they may be in both physico-chemical make-up and behaviour, and whatever the achievements of skilled engineers in the future.

In Part 2 of this paper I shall indicate what may be called the 'traditional' answer to the problem. I shall then state what I take to be the important objection to it. In Part 3 I shall formulate what I shall call the 'verificationist' approach to the problem. I have no single philosopher in mind here, but I use the label 'verificationist' to indicate a general line of approach, one of whose characteristics is to be specially concerned with problems of verification. I shall indicate that this approach, too, is not without its difficulties. In Part 4 I shall offer my own contribution, which makes use of the concept of 'body-schema', and provides, I hope, some sort of synthesis between the conflicting views of 'traditionalism' and 'verificationism', as well as

¹ This, of course, is not the same as asking whether a second-generation *homo mechanisma* *would in fact* be a man or a machine—a question on which it scarcely seems profitable to speculate.

avoiding the difficulties inherent in each. In Part 5 I shall make some concluding remarks on this paper as a whole.

2 The 'Traditional' Answer

According to the 'traditional' answer, the cardinal difference between men and machines is that men have *minds* and are *conscious*. However complex the responses which a machine may produce, it is still not a man unless at some stage it is actually conscious. If the second-generation homo mechanisma has no mind and does not become conscious, then it is a machine and not a man.

Criticism. It might be argued that this answer suggests by implication that *having a mind* is somehow like *having a camera* or *having a violin*. Because the words 'camera' and 'violin' refer to *things*, there is danger in supposing that the word 'mind' refers to a *thing* also—in this case a *recondite* and 'immaterial' thing.

The argument here is on strong ground. Clearly there is no means of telling whether such an entity is present or not. It is what Ryle calls a 'ghost-in-the-machine';¹ it has no visual appearance, no weight, no taste, and no smell, nor can it have any effect on any recording device. To postulate the existence of such an entity is to formulate an unverifiable hypothesis—unverifiable, not as a result of practical difficulties, but unverifiable because there is not and cannot be any account of what would constitute verification. Such an entity must remain forever unknown and unknowable, and its presence or absence cannot therefore be the criterion by which we distinguish men from machines.

3 The 'Verificationist' Answer

The 'verificationist' answer can be formulated by saying that there *could not be* grounds, other than those of ancestry, for distinguishing a second-generation homo mechanisma from an ordinary human being.²

¹ G. Ryle, *The Concept of Mind*, London, 1949, pp. 15–16

² An interesting formulation of this viewpoint occurs in A. J. Ayer, *Language, Truth, and Logic*, London, 1949, p. 130. Ayer writes: 'When I assert that an object is conscious I am asserting no more than that it would, in response to any conceivable test, exhibit the empirical manifestations of consciousness. I am not making a metaphysical postulate concerning the occurrence of events which I could not, even in principle, observe.'

ON DIFFERENCE BETWEEN MEN AND MACHINES

This conclusion is inevitable, it would be said, simply from the way in which the problem has been stated. 'Homo mechanisma' has been defined in such a way that his physico-chemical make-up and his behaviour are assumed to be no different from those of true human beings. If, however, this assumption is made, and if, further, it is agreed to ignore distinguishing criteria in terms of ancestry, then what we are doing is to ask for distinguishing criteria with one hand while with the other we make sure that no distinguishing criteria are possible.

On this view it is futile, of course, to suppose that the difference between men and machines is that men are occupied by a recondite 'ghost' and that machines are not. The important differences, it would be said, between men and machines are those cited in Part I—machines look different, are made differently, and except in a few respects behave differently from human beings. If these and similar differences are removed, then *ex hypothesi* no differences are left.

It is still legitimate, on this showing, to say that, as things are, men are conscious and machines are not, provided the word 'conscious' is interpreted behaviouristically. To tell if a person is conscious, it might be said, we study his *behaviour*. There are all sorts of ways of deciding if a person is conscious—the opening of the eyes, the responses to touching and other stimuli, the making of purposeful movements, and so on. A surgeon or an anaesthetist who failed to use such criteria would soon be out of a job. But, as a defender of the 'verificationist' view would point out, these distinguishing criteria are not relevant to the problem as formulated, for whatever criteria are given of this behaviourist kind, a machine behaving in exactly similar ways has been assumed to be a possibility.

In the case of homo mechanisma, it would be added, the only things that can profitably be discussed are questions of policy. People might wonder, for instance, whether homines mechanismata should be given the vote when they reach the age of twenty-one, whether it would be sensible to threaten them with hell-fire, and whether they could have *mens rea* in the law courts.¹ What is called for is not examination of the past history of a homo mechanisma but a decision of policy. As Wittgenstein has somewhere put the matter, 'What does or does not constitute a cow is for the public to decide'.² If we

¹ See Justinian, Institutes III, 107: 'Non reus nisi mens sit rea'

² Compare also the discussion by A. G. N. Flew, of 'Is a flying-boat a ship?', *Logic and Language*, Oxford, 1951, p. 3

want to give homines mechanismata the vote, etc., we *decide to call* them men, not machines. To rule that they have 'souls' is to rule that they shall be treated in certain ways. Objects without 'souls' are by definition objects which cannot achieve or forfeit salvation.

Finally, it would be said that if homines mechanismata and their descendants became very numerous, if 'machine-origin' was no barrier to promotion, and if ordinary machines—such as present-day calculating machines—were no longer used, then there would be no need for the separate words 'man' and 'machine'. One word would do for both, and the problem of finding criteria for distinguishing them would not arise. Until that blessed, or otherwise, state of affairs arises, however, both words are needed; and at present, in the absence of machines that really do behave like men, it is perfectly easy to tell the difference.

Criticism. The strength of the 'verificationist' answer is its resolute refusal to entertain unverifiable hypotheses. Its weakness is the weakness of any theory that regards 'behaviour' as the important concept rather than 'consciousness'. Let us now look at this weakness in more detail.

First, it is difficult not to be suspicious of the argument, quoted above, about the anaesthetist who uses behavioural criteria to tell if his patient is conscious. We all know that there is a difference between watching someone showing signs of pain and feeling pain oneself. In the same way it seems indisputable that there is a difference between watching someone showing signs of consciousness and being conscious oneself. Now in the 'verificationist' view, 'X is showing all possible signs of consciousness but is not actually conscious' would be self-contradictory; the phrase 'being conscious' can on the 'verificationist' view be 'unpacked' into 'showing all the signs of consciousness'. Yet it seems clear that there is no self-contradiction here. We may be puzzled as to how anyone could ever know the truth of 'X is showing all possible signs of consciousness but is not actually conscious', and it might even be argued that on no conceivable occasion would anyone have the right to say it. But it is clearly not self-contradictory in the way in which 'This object is behaving in all possible respects like a telephone but is not a telephone' is self-contradictory. The 'verificationist' view is forced to assume a parallelism which clearly does not hold.

Secondly, though the 'verificationist' view warns us that the difference between men and machines is not the presence in the former

of a recondite 'ghost', it gives no adequate account of how the idea of a ghost-in-a-machine could ever have arisen. It is hard to see how this could have happened if there were not *something* in our experience to suggest it. As I shall argue later in the paper, the idea of a ghost-in-a-machine is not really quite so stupid a blunder as Ryle's criticism might lead us to suppose.

What is required, therefore, is a theory which both avoids these difficulties in the 'verificationist' answer and is also able to meet the objection to the 'traditional' answer that it involves an unverifiable hypothesis. I propose now to offer a theory which seems to me to satisfy these requirements.

4 *The Concept of 'Body-schema'; a Suggested Synthesis*

Before I present my central thesis, a brief introduction to the concept of 'body-schema' is called for.

The career of this concept has been somewhat chequered.¹ Head,² who introduced it, used it chiefly as an explanation of appreciation of changes of posture; Bartlett³ and Oldfield⁴ have used it as an improvement on the earlier 'trace' theory of memory. Head, J. O. Wisdom,⁵ and others have indicated its relevance for the phenomenon of phantom-limbs. To what extent my use of the term in this paper coincides with its use by these other writers is unimportant; and indeed there seem to be no very explicit rules as to how we should operate with it. Oldfield and Zangwill make clear that associations with Kant's use of the word are misleading, and that, if there is a parallel with Kant, it is with his *pure intuitions* of space and time in the Transcendental Aesthetic.⁶ It is not clear from existing literature whether we can ask what a schema looks like, feels like, smells like, what is its size, weight, or age, whether people

¹ For a history of it, see R. C. Oldfield and O. L. Zangwill, 'Head's Concept of the schema and its application in contemporary British psychology', *The British Journal of Psychology*, 1942, 32, and 1943, 33. Also, for references, and a brief but valuable discussion, see W. R. Brain, *Mind, Perception and Science*, Oxford, 1951, p. 17.

² H. Head, *Studies in Neurology*, London, 1920, Vol. II, pp. 605-608, 722-726

³ F. C. Bartlett, *Remembering*, Cambridge, 1950

⁴ R. C. Oldfield, 'Memory mechanisms and the theory of schemata', *The British Journal of Psychology*, 1954, 45

⁵ J. O. Wisdom, 'The Concept of Phantom-body' *Actes du xième congrès international de philosophie*, 1953, 7

⁶ I. Kant, *Critique of Pure Reason*

can possess one without knowing it, nor whether they can think they possess one and yet be mistaken ; and, if some of these questions can be asked and not others, it is not clear why this should be so. Indeed Bartlett's phrase 'turning round on our schemata' seems to me so obscure as to be unintelligible. In view of this uncertainty I propose to give a 'stipulative' as opposed to a 'lexical' account of the word. That is, I shall *stipulate* how I wish the word to be understood rather than examine existing literature in the hope of finding a standard or correct usage.¹

For purposes of this paper there are two special points with which I am concerned—abnormal clinical findings and our awareness of ourselves.

The clinical findings in question are of two kinds—the occurrence of phantom-limbs and the inability of some patients with a cortical lesion to appreciate whereabouts on their body they have been touched. Both these phenomena are so well known that detailed description is unnecessary. The relevance of the concept 'schema' is as follows. The brain, it seems, must contain some symbolic representation of the structure of the body. Only so, in normal people, can, say, pressure on a particular finger be appreciated as coming from that finger. Some cortical structure must be involved which allows for comparison with other parts of the body. As a result of cortical damage the ability to make this comparison may be impaired. Again, if the brain is intact but a limb has been amputated, the conditions for awareness of the missing limb may still be present in the brain, and the person is therefore aware of a phantom-limb. Head indicates the link between the two types of occurrence as follows :

One of our patients had lost his left leg some time before the appearance of the cerebral lesion which destroyed his power of recognizing posture. After the amputation, as in so many similar cases, he experienced movements in a phantom foot and leg. But these ceased immediately on the occurrence of the cerebral lesion ; the stroke which abolished all recognition of posture destroyed at the same time the phantom limb.

What is postulated is in effect a cortical structure which serves to explain both phantom-limbs and ability to appreciate changes of posture.

The word 'schema', I suggest, should be used to refer not to this structure itself nor to a person's actual body, but to what may be

¹ For a more detailed account of the difference between lexical and stipulative definition, see R. Robinson, *Definition*, Oxford, 1950

called the *phenomenological counterpart* of this body. Wisdom calls it the 'phantom-body'. He writes, 'Let us now extrapolate the idea of a phantom-limb and invent the concept of a phantom-body. . . . It would be, as it were, a ghostly filling that occupies the physiological body.' The word 'phenomenological' is formidable, and to many it is even suspect. I understand the word 'phenomenology' to mean 'the study of the way in which things appear'. The notion of a phenomenological counterpart to our ordinary body (or, as Wisdom calls it, our 'physiological' body—the body that the physiologist investigates) is legitimate provided it indicates the way things do in fact feel. That they do feel like that seems to me beyond any doubt.

This brings us to the question of normal self-awareness. The fact that the cases so far cited are abnormal and known to be abnormal must surely indicate that we know what normal awareness of our body-schema is like. Some of our uses of the word 'I' and 'person' are relevant here. Wisdom writes: 'Sometimes the pronouns "I", "we" and so on, which denote a person, refer to the phantom-body. . . . Should we then equate the phantom-body, not with the person, but with the mind?' Wisdom does not commit himself fully, but many other writers have made suggestions that are largely in conformity with such an idea. Thus Berkeley¹ speaks in the same breath of 'mind, soul, spirit, or myself'. Koffka² writes:

The local kinaesthetic processes, since in their entirety they help to organise the ego, are not independent local events, but part events in a larger system of events. If then a visual datum is welded with a kinaesthetic one, it must also of necessity become a part in a greater whole, i.e. it must be incorporated in the Ego system.

Scott³ has called attention to the connection of 'body-scheme' with 'ego'.

The point can be brought out further if we consider the question of the *limits* of our body-schema. Isaacs⁴ says that in the case of the very young child 'the skin is not yet felt as the boundary between

¹ G. Berkeley, *Principles of Human Knowledge*, section 2

² K. Koffka, *Principles of Gestalt Psychology*, New York, 1935, p. 329

³ W. Clifford M. Scott, 'Some Embriological, Neurological, Psychiatric, and Psycho-analytic Implications of the Body-Scheme', *International Journal of Psycho-analysis*, 1948, 29, Part iii

⁴ S. Isaacs, 'The Nature and Function of Phantasy', *International Journal of Psycho-analysis*, 1948, 29, Part ii

inner and outer reality'. Flügel¹ says 'Our clothes form an intermediate layer between ourselves and the outside world'. Ritchie² writes, 'When I look at something what I see is outside my body. . . . The pain from a blister in my heel is inside my body. What I can see on the heel or touch with my finger is not quite outside my body but is the outside of it.' Ryle³ quotes the examples, 'I was not scorched; only my hair was' and 'I collided with the pillar-box'. In the one case it seems that our hair is not part of our body-schema, in the other case that our body-schema extends to the limits of the car which we are driving. Perhaps the area which we can *control* is relevant. A tennis racquet, properly controlled, becomes, as people say, 'part of' the player.

Enough has been said, I think, to indicate how I wish the word 'schema' to be understood for present purposes. A few words should now be said on the notion of *externality* and *internality*. There is a straightforward sense of these words, in which we speak, for instance, of an external or internal coal-shed—that is, a coal-shed inside or outside a house. In the same way we could speak of muscles *in* my leg, or processes *in* my brain. In all these cases there is the possibility of measurement, and what is said to be inside or outside is visible, tangible, and solid. To stress the connection with measurability I shall speak of *dimensional* externality and internality. When, however, we use phrases such as 'the external world' or 'external reality', what is meant is something *external to me*, or, as I should prefer to put the matter, *schematically*⁴ external, external in relation to the body-schema. In this case, what is external is visible, tangible, and solid, just as before, but the same does not hold of what is internal. Internal to the body-schema are such things as pains, aches, and tickles, and also mental images and those diffuse, vaguely localised feelings which we appreciate as part of our 'inner life'. It makes no sense to speak of seeing and touching such things,—at any rate not in the way in which we see and touch chairs and tables. Moreover, there is no question of exact quantitative measurement. There is a pseudo-quantitative assessment of distances. I use the word 'pseudo' to indicate that no measuring device can ever check such assessment.

¹ J. C. Flügel, *The psychology of clothes*, Hogarth, 1950

² A. D. Ritchie, *The Natural History of Mind*, London, 1936, p.149

³ Ryle, *op. cit.*, p. 189

⁴ Some may prefer 'phenomenologically external'. This term was suggested to me by Dr J. O. Wisdom.

ON DIFFERENCE BETWEEN MEN AND MACHINES

I might estimate, perhaps, that tickle B felt about twice as far from tickle A as did tickle C ; but no one can put a ruler alongside my tickles and say that I am wrong. (At best he could put the ruler *where I told him* I felt the tickle, but this is quite different.) Thus, while a muscle in my leg is dimensionally internal a pain in my leg is schematically internal. One of the functions of the word 'in' is to call attention to certain sorts of positional relationship. Thus an image in a mirror is positionally localised in relation to the mirror-frame. The spatial relationship between a tickle and the 'place' where I take my leg to be is in part like, in part unlike that between a muscle and my actual leg or that between mirror-images and mirror frames.

It may be objected that to postulate the existence of a body-schema and to introduce the schematic sense of external and internal is to be misled by a metaphor. The pain is not in my actual leg ; and the objection would be that we are therefore postulating a metaphorical non-actual 'phantom'-leg or 'phantom'-body simply as somewhere to house the pain. This objection raises complex questions connected with the rules for using such words as 'actual', 'real', and 'exists'. Here is no place to discuss even what could be *meant* by arguing whether our ordinary body or our 'phantom-body' is 'the real one'. I entirely agree that we do not literally possess *two* bodies, one of each sort, nor even *two things*, a mind and a body. Instead of 'we are literally aware of a phantom-body' I am quite prepared to substitute 'we are phantom-ly aware of an ordinary body', if this is thought to save us from asking impossible questions about the ontological status of phantoms. What is important for present purposes is to mark the distinction between the schematic and dimensional senses of 'external' and 'internal' and to show that the former involves a relationship to what we think of as ourselves.

We are now in a position to put forward our own account of the difference between a man and a machine. The 'traditional' view, that man is *conscious* and has a *mind* or *soul*, is not so far wrong after all. On the present view this amounts to saying that a man, unlike a machine, *possesses awareness in relation to his body-schema*. In other words he can recognise pains, aches, and tickles as internal, and can distinguish them from houses, trees, and other objects which are perceived as external. ¹

¹ This is to say, in effect, that a man, unlike a machine, would be aware of spatial intervals. The same also holds of temporal intervals ; but discussion of this question

Whether a second-generation homo mechanisma would possess such awareness we just do not know. It is possible that any artificially made contraption might do so, even if it was very unlike a human being in appearance. All that is being suggested is that *if* a homo mechanisma possessed this awareness, then it should be labelled 'man' not 'machine'; the presence of this awareness would be sufficient to discount the assertion that it was a machine.

It remains to show how this viewpoint meets the objections levelled at the 'traditional' and 'verificationist' views.

The first objection against the 'verificationist' view was that it forced us to say that 'X is showing all possible signs of consciousness but is not actually conscious' was self-contradictory, whereas it seems clear that, whatever the oddity of this sentence, it does not actually contradict itself. Now, if to be conscious is to have awareness in relation to the body-schema, the difficulty disappears. There *is* a difference between showing *signs* of this awareness and actually having it. *Ex hypothesi* a homo mechanisma shows all the signs that could be asked for. It does not follow that a homo mechanisma is conscious. It might be; but even if it were, this is not a matter of logical necessity.

The second objection against the 'verificationist' view was that it gave no account of how 'ghost-in-machine' language could ever have arisen. Now we may certainly agree that to think of a ghost as *dimensionally* inside a machine is grotesque and misleading. It suggests that ghost and machine are of comparable 'ontological status', in the sense of being part of the 'stuff' out of which the universe is made; and we may well be puzzled as to how such a ghost could be an object of knowledge at all. But the phrase 'ghost-in-the-machine' loses much of its power of ridicule if we think in terms of schematic, not dimensional, internality. If the idea of a body-schema or 'phantom-body'¹ is at all helpful, then 'ghost-in-machine' language is not just the curious aberration which Ryle seems to suggest.

The third requirement of our theory was that, unlike the 'traditional' theory, it should not involve an unverifiable hypothesis. Now it may be objected that this is precisely where the present theory

would take us beyond the scope of the present paper. Compare Kant, *op. cit.* Kant's point could perhaps be reformulated in our present terminology by saying that, without reference to a body-schema, things could not appear as spatial or temporal at all.

¹ Compare Wisdom, *op. cit.*

breaks down, and that what is being postulated is an awareness ('awareness in relation to the body-schema') which is unknown and unknowable and whose presence could not be detected by any conceivable criteria.

Let us look at this objection in more detail. The attack on 'unverifiable hypotheses' has been formulated in various ways. Thus, according to the views formerly expressed by Bridgman,¹ 'X has awareness in relation to a body-schema' would be vulnerable because it contains a collection of words which do not admit of 'operational definition': that is, there can be no account of what *operations* we would have to perform to tell if the sentence was true or false, or, more strictly, the sentence cannot be re-defined in terms of a set of operations. The words cannot be given 'cash-value', to use Price's phrase,² in terms of observable occurrences. To put the matter another way, they are vulnerable under the *verification principle* as expounded by Ayer³ and others. These formulations of the difficulty differ from one another to some extent, but for our purposes these differences can be ignored.

The objection can, I think, be met by consideration of what exactly is being demanded. If someone says, 'How can I ever know the truth of "X has awareness in relation to a body-schema"?' his demand can only be for external criteria—criteria, that is, which would be schematically external to any observer. This demand is perfectly sensible as far as it goes. We can tell to some extent by external criteria whether a person has normal awareness of his body-schema. There are grounds for supposing, for instance, that very young children and some mentally ill patients do not have it. Again, if a person showed no more signs of pain at losing a finger than he did as a result of a hair-cut, we would certainly suspect that there was something odd about his body-schema. Possibly it is a disturbance of this sort that leads certain individuals to be 'accident-prone' and to catch their fingers in machinery. What is clear is that a demand for external criteria can at best give us *signs* of awareness rather than *samples* of it, in much the same way as footprints in the flower-beds are signs of the presence of a burglar but not themselves samples of what is meant by 'burglary'. The whole matter is no more and no

¹ P. W. Bridgman, *The Logic of Modern Physics*, New York, 1927

² H. H. Price, 'Image Thinking', *Proceedings of the Aristotelian Society*, 1951-52, 52, 152

³ Ayer, *op. cit.*, chapter I

less puzzling than our inability to know for certain if animals feel pain or if our fellow-humans are lying when they tell us that they have toothache. If behaviour, including 'verbal' behaviour, is displayed suggestive of toothache and there is no behaviour that indicates differently, then we have 'all the reasons we could have'¹ for saying that a person has toothache; and if an animal shows a number of different signs of pain, we have 'all the reasons we could have' for saying that it is in pain. In short, if we ask for schematically *external* criteria for deciding something and then lament because what is schematically *internal* is not forthcoming, our lamentation is clearly misplaced. 'X has awareness in relation to a body-schema' is 'uncashable' in just the same way as 'X has toothache' or 'Animals can feel pain' is 'uncashable'. The fault lies, not in these sentences themselves, but in the sort of 'cash' that the objector is expecting.

We may conclude, therefore, that the protest of 'unverifiable hypothesis' cannot be made the basis of an objection to our theory.

One final comment may be made in connection with the general subject matter of this paper. An engineer who had constructed a homo mechanisma would *ex hypothesi* know every detail of how the human brain functions. But it would still make no sense to *tell him to create* an object which could relate to a body-schema. The most he could be told to do would be to create the conditions under which this ability occurred; he would then simply have to hope that the ability would supervene. Again, if he were told to construct a mechanism for converting familiar objects such as tables into the symbolic representation of these objects in the cortex, he would be happy to comply. But, however great his competence, it would be muddled thinking to tell him to make a comparable mechanism for converting patterns in the cortex into awareness in relation to the body-schema. It is the alleged impossibility of creating such a mechanism which is often regarded as one of the most baffling features of the so-called 'mind-body' problem. On the one hand awareness in relation to the body-schema appears to require explanation; on the other hand nothing that we could observe would ever give us the sort of explanation that we seem to want.

If the general argument of this paper is correct, the mistake lies in speaking as if awareness in relation to the body-schema were somehow a *link in a chain*. Thus Kapp, for instance, writes :

¹ Compare John Wisdom, *Other Minds*, Oxford, 1952, p. 232

ON DIFFERENCE BETWEEN MEN AND MACHINES

There are the afferent nerves that bring messages from the eye to the brain, a multitude of synapses in the brain itself, efferent nerves that carry controlling impulses from the brain to the operator's finger. To avoid vagueness one must discover exactly *at which of the many links the act of observation occurs*.¹

Kapp speaks also of *locating* such an act. This seems to me part of the same mistake. To locate something is to discover where in space it is, and perhaps, by analogy, to discover when in time it occurred. But no measuring device could ever tell us that an act of observation was four feet to the left of the piano, and no act of observation can either succeed or fail in setting a timing device in motion. Kapp is speaking throughout as though X's awareness in relation to the body-schema could somehow form part of Y's external world. What we need to realise is that, whatever a neurologist discovers, and however much he improves his techniques of observation, *he is always studying what is schematically external to himself*. He may succeed as much as you please in explaining one external event in terms of another ; someone else's awareness in relation to the body-schema can never be part of what he discovers. Adrian² seems in effect to recognise this when he says, 'The real trouble comes from the feeling that there may be an important part of the picture which can never be fitted in however long we may look at it'. This language suggests, however, by implication, that there is still something which *ought* to be fitted in. If the present argument is right, the whole idea of a mysterious and inexplicable 'gap' in our knowledge is mistaken, and it makes no sense either to lament the existence of such a gap or to express the hope that one day we shall discover the required information. Gaps in the neurologist's knowledge can be filled only by discovering events that are schematically external.

5 Conclusions

No attempt has been made in this paper to surmise on what developments are likely to take place in the fields of electronics, neurology, and elsewhere, nor to contribute anything towards the problems of constructing a homo mechanisma. I have been concerned

¹ R. O. Kapp, 'The Observer, the Interpreter, and the Object Observed', *Methodos*, 1955, 7, 4

² E. D. Adrian, *The Physical Basis of Mind*, Ed. Laslett, Oxford, 1950, Chapter 2, p. 5

to discuss how the concepts 'man', 'machine', 'mind' and 'consciousness' would operate if it ever happened that a homo mechanisma were constructed. Such a discussion serves to bring out how these concepts function at the moment. As things are, men differ from machines, (a) in appearance, physical make-up, and behaviour, and (b) in having awareness in relation to the body-schema, i.e. in being able to recognise houses and trees as external to the body-schema, and pains, aches, and tickles as internal. This paper is concerned to point out that, even if all distinguishing marks of the (a) variety were removed, as is at least possible, there is still the difference mentioned under (b). Men can relate their perceptions to a body-schema, but in the case of machines there is no body-schema to which perceptions could be related.

University College
Bangor

EXPLANATION, PREDICTION, AND ABSTRACTION *

ISRAEL SCHEFFLER

Introduction

IN recent philosophy of science, three basic views concerning explanation and prediction have received wide support, attaining almost canonical status. They are (A) the view that explanation and prediction share a *common structure*, with but the pragmatic difference that an explained event antedates the statement of its explanation while a predicted event can only follow its prediction, (B) the view that explanation and prediction represent the *central purpose* of science and are epistemologically basic, and (C) the view that explanation and prediction are abstract in reference, their objects being not concrete things but *idealistic or intensional entities* like phenomena, facts, or states-of-affairs. I shall argue, in what follows, that these three views are untenable, and I shall affirm instead that explanation and prediction are structurally distinct, that, associated with control, they are subsidiary to the primary concern of science with comprehensive relationships among events, and that they require no abstract, idealistic entities as objects.

I *The Structural Identity Claim*

The notion of explanation has a variety of uses both in ordinary speech and in scientific contexts. In both spheres we speak alternatively of explaining concepts or terms, laws or generalisations, and concrete occurrences or events. It is interesting to observe, at the outset, how this very variety of uses contrasts with those of the notion 'prediction'. For while we speak of predicting occurrences or events, we surely do not speak of predicting concepts or terms, nor, in any obvious sense, laws or generalisations. The claim of structural identity is made, however, not with reference to all patterns of scientific

* Received 4. viii. 55. The writer wishes to thank Professors C. G. Hempel, N. Goodman, W. V. Quine, and N. Chomsky for critical comments.

explanation, but specifically regarding explanation of events, usually described somewhat as follows.¹

Let a and b be distinct events, described by the sentences A and B respectively, and let L be a law or conjunction of laws. Suppose, also, that B is a logical consequence of A and L , but not of A alone. If A and L are true, while b has already occurred, we may say that b has been accounted for or explained by the conjunction of A and L , or that this conjunction satisfies the requirements for an explanans of the explanandum B .

This sketch incorporates the four conditions listed by Hempel and Oppenheim in their discussion of the logic of explanation,² which we take here as a model :

- (R1) The explanandum must be a logical consequence of the explanans.
- (R2) The explanans must contain general laws required for the derivation of the explanandum.³
- (R3) The explanans must have empirical content.
- (R4) The sentences constituting the explanans must be true.

Justifying (R4) as contrasted with an alternative requirement of high confirmation for the explanans, Hempel and Oppenheim cite the case of a purported explanans highly confirmed at time t_1 but later highly disconfirmed at time t_2 , in which event we should not wish to say that what was an explanans at t_1 ceased to be one at t_2 but should rather prefer to assert that, while its truth had been probable relative to available evidence at t_1 , its falsity was probable at t_2 , and correlatively, its inadequacy as an explanation at any time.

It is this pattern which is alleged to be identical with that of scientific prediction. As Hempel and Oppenheim put it, 'the same formal

¹ The description here discussed (as well as the structural identity claim) is given in various forms by a number of authors. Among recent empiricist writings, the following should be especially mentioned : K. R. Popper, *Logik der Forschung*, Wien, 1935, pp. 26 ff., and *The Open Society and its Enemies*, London (first published 1945) 1947, Vol. II, p. 249 and pp. 342-343, C. G. Hempel, 'The Function of General Laws in History', *Journal of Philosophy*, 1942, 39, 35-48, and C. G. Hempel and P. Oppenheim, 'Studies in the Logic of Explanation', *Philosophy of Science*, 1948, 15, 135-175.

² Hempel & Oppenheim, op. cit. pp. 137-138

³ Hempel and Oppenheim do not also require the explanans to include one statement which is not a law, for 'to mention one reason', they wish to consider an explanation of generalisations a *bona fide* explanation. In the light of what was observed above, regarding the restriction of the notion of 'prediction' to events, it would seem unlikely that they wish to extend their statement of the explanatory pattern to the prediction of generalisations, though they do not make this point explicit.

EXPLANATION, PREDICTION, ABSTRACTION

analysis, including the four necessary conditions, applies to scientific prediction as well as to explanation'. The pragmatic difference as they, in agreement with several other authors, formulate it consists in the fact that for explanation, *B* is given, *b* having occurred, and the conjunction of *A* and *L* is provided afterwards, while for prediction, this conjunction is given and *B* is derived prior to the occurrence of *b*. Explanation, as they put it, 'is directed towards past occurrences', prediction 'towards future ones'.

Now this account implies that every explanation, if stated prior to the event described by its explanandum, would be predictive, while every prediction, stated after the event in question, would be explanatory. The first of these consequences is indeed explicitly drawn in the study of Hempel and Oppenheim, 'It may be said, therefore, that an explanation is not fully adequate unless its explanans, if taken account of in time, could have served as a basis for predicting the phenomenon under consideration.' The second consequence, though equally necessary for the structural identity in question, is not further elaborated.¹ In view of the following considerations, it seems to me that it is untenable.

(a) First, note that 'is a prediction' is not properly applicable to abstract sentences or propositions, since the same sentence 'It rains on May 8, 1952' is or is not a prediction depending on the temporal circumstances of its utterance. Or, more accurately, since what predicts must have an appropriate time relative to what is predicted, and since abstract sentences or propositions are non-temporal altogether, they cannot be properly denoted by 'is a prediction'. We may distinguish the abstract sentence from the uses made of it at various times and denote uses as predictions, if we like. Alternatively, we shall here construe 'is a prediction' as predicable of concrete utterances or inscriptions (i.e. tokens) with temporal boundaries, but the point to be made can be readily put in terms of the other usual analyses.

Consider now any utterance or inscription of declarative, non-compound form. In accordance with the dominant ordinary notion of prediction, any such utterance or inscription is a prediction if it explicitly asserts something about some time later than any of its own.

¹ This sentence does, however, appear (p. 138) though it does not figure importantly in the later treatment of the authors: 'only to the extent that we are able to explain empirical facts can we attain the major objective of scientific research, namely not merely to record the phenomena of our experience, but to learn from them, by basing upon them theoretical generalisations which enable us to anticipate new occurrences and to control, at least to some extent, the changes in our environment.'

But it is clearly false that restating each such prediction following the time of its predicted occurrence, explains this occurrence even when both prediction and restatement are true. Thus, no inscription like 'Eisenhower is elected President on November 4, 1952' *explains* Eisenhower's election, though every such inscription prior to November 4, 1952 is a prediction in the ordinary sense, and true at that. The point then is that, in the usual sense of 'prediction', not every restatement of a prediction after the event is explanatory, even though every statement of an explanation prior to the event is predictive.

Nor will it do to invoke epistemology at this point by asking how a prediction of Eisenhower's election could have been made without the use of general laws and statements of relevant antecedent conditions. That the methodological genesis of a prediction does not meet rational or scientific requirements may involve irrational behaviour by its producer, but is no bar to its predictiveness in the ordinary sense, nor even to its truth. Clairvoyants, prophets, and news-commentators all predict in the sense under consideration, just as do scientists. For pragmatists and positivists in particular, who justify scientific method by success in prediction, a restriction of the latter to *scientific* prediction would reduce their justification to triviality.

(b) Suppose, however, that structural identity is interpreted as holding between explanation and *rational* prediction as *practised in the sciences*, involving reference to general statements and specific condition statements.¹ Even so, it will appear that not all predictive restatements

¹ One might, incidentally, raise the question whether scientific prediction always takes the form mentioned, in view of predictions only inductively well-grounded, but not implied by any relevant conjunction of universal generalisations and condition statements. For example, suppose we predict that the 5,000th ball drawn at random from an urn will be red, since all have heretofore been red with, say, the exception of the first drawn. If such prediction is acknowledged as rational, surely its full restatement after the event is non-explanatory even if true. Thus, suppose our prediction is fulfilled and we are told that the explanation why the 5,000th ball was red is because the previous 4998 were. This seems unacceptable as an explanation though it restates the rational grounds for having made the prediction previously. Indeed, were there not a single exception, the fact that, e.g. 4,999 balls were red does not explain why the 5,000th is red though it does rationally ground the prediction that it will be. This example serves to illustrate a distinction, of some general importance, between asking 'Why P?' in the sense, 'What is an explanans for P?' and in the often divergent sense, 'What are rational grounds for asserting P?' The distinction is sufficient to invalidate a widespread identification of the *explanation of A* with *showing that A was to be expected*, as suggested, e.g. by Toulmin, *The Place of Reason in Ethics*, Cambridge, 1950, pp. 122 ff., and many others.

EXPLANATION, PREDICTION, ABSTRACTION

after the event are explanatory, since predictive success involves the possibility of predictive failure, i.e. false predictions. But no explanation is false, since it consists of an explanans, which by (R₄) must be true, and an explanandum which, being a logical consequence of the latter, cannot be false either. This divergence is related to the use of scientific predictions in testing the body of assumptions at a given time ; for such testing to occur, it must make sense to judge a derived prediction false, thereby forcing a revision in its ground-premises. Indeed, to the extent that predictive test is involved in confirming the truth of general laws, themselves required for explanation by (R₂), to that extent the possibility of falsifying predictions is presupposed by the confirmation of explanations.

If this divergence between scientific explanation and prediction is granted, one might still attempt to reinterpret the structural identity claim as holding between rational prediction (by deduction from general statements and condition statements) and *proffered* explanation, which, of course, may be false. Without artificial restriction of the latter notion, however, such reinterpretation fails. For in its ordinary sense, 'proffered explanation' refers not only to certain explanations which are false, i.e. violate (R₄), but also to some which fail to exhibit the required logical character as specified by (R₁), fail to contain general laws as required by (R₂), or lack the empirical content demanded by (R₃). Artificial expansion of this notion would, moreover, also be necessary since we would not, ordinarily, consider any account to be even a proffered explanation unless it purported to explain some *fact*, i.e. unless at least its explanandum were true ; derived *predictions*, on the other hand, may clearly be false. To specify then explicitly that we are to require fulfilment of (R₁), (R₂), and (R₃), but neither (R₄), nor the truth of the explanandum, (while certainly legitimate and often convenient for other purposes), renders trivial the claim of structural identity between prediction and explanation. For, if true, this claim appears no longer a surprising description of two antecedently-known patterns which happen to correspond, but rather a consequence of our deliberate theoretical tampering. What is finally correlated to the term 'prediction' is a technical artifact independently related neither to the ordinary sense of 'explanation' nor to that of 'proffered explanation'.¹

¹ Furthermore, it is doubtful if even this explicit delimitation of just what is supposed to correspond to prediction is sufficient. For on one widely-held view (shared by Hempel) abstract, partially interpreted theories are an integral part of

We conclude, then, that the structural identity claim should be rejected ; far from differing only in pragmatic relationships, explanation and prediction have different logical characteristics : explanations are true, predictions need not be ; making predictions is part of one way of confirming the existence of explanations ; predictions may be made with or without rational grounds, and some rational grounds adequate for prediction fail to *explain* the predicted occurrences. If, then, the structural identity claim is to be approved for implying that every explanation must have been capable of prediction, it is no less censurable for glossing over these important differences in structure and rôle.

2 *The Centrality Claim*

If explanation and prediction are structurally distinct, the usual claim of centrality in scientific procedure is at least ambiguous : are they both central, or is one more important than the other ? The following considerations will specify some distinctive temporal asymmetries peculiar to each, arguing for their irrelevance to general scientific inference, and hence the inadvisability of incorporating them into typical or central models of such inference.

(a) Consider first prediction. A necessary condition for the predictive character of an utterance or inscription is its asserting something about some time later than its own. This is the force of future-tense indicators often taken as a sign of predictive character in ordinary usage, though, of course, not essential to such character. If, now, we examine the four requirements of Hempel and Oppenheim, we find no temporal conditions among them.¹ To be sure, this pattern of requirements is intended to reflect the inferential process of making (scientific) predictions, but the inferences admitted by the pattern include other types as well, which cannot be classed as explanations either. Thus, even if this pattern is not exhaustive, but represents

advanced sciences and, as such, are essential to rational prediction in those sciences. It is unclear, however, in what sense we may speak of the truth or falsity of *partially interpreted* formal systems. If we cannot, then such theories, though predictive, are not even proffered explanantia in our explicit sense. For though this sense does not *require* truth, it does require that the explanans be, in point of fact, either true or false.

¹ Often, indeed, terms like 'consequence', 'derivation', 'antecedent condition', 'presuppose', etc., are ambiguously employed with occasional temporal reference. It does not appear, however, that such interpretation is here intended.

EXPLANATION, PREDICTION, ABSTRACTION

one scheme of scientific inference, it is much wider and more general than inferences of predictive or explanatory nature.

Thus, as the distinction between explanation and prediction is drawn by Hempel, Oppenheim, and others, it would be said, with reference to our earlier example, that if *B* is given, i.e. if we know that the phenomenon described by *B* has occurred, and a suitable set of statements *A* and *L* is provided afterwards, we have an explanation, while if the latter statements are given, and *B* is derived prior to the occurrence of the phenomenon it describes, we speak of a prediction. Note, however, one way in which this description fails to exhaust the inferences allowable by the pattern: If *A* and *L* are given rather than *B*, thus precluding explanation, their logical consequence *B* may be derived not prior to, but simultaneous with or after the occurrence of *b*. For example, *b* may have occurred prior to *B*'s derivation but later than *a*, or it may have occurred prior to *a*.¹

For an illustration of the first case (i), consider an astronomer who, from statements describing the *beginning of an ancient eclipse a*, plus the appropriate laws, deduces a statement describing *its end b*. For an example of the second case, (ii) imagine the same astronomer who, from appropriate laws plus statements describing *some relevant configuration of heavenly bodies at some time during his own personal experience a*, deduces statements describing *some eclipse in former times b*. In neither case do we have a prediction, yet both inferences fulfil the pattern in question.² What is common to both and to the predictive inference in question is not any temporal relation between statements and described events, but rather the givenness of *A* and *L* and the later derivation of *B*. The latter sequence, however, bears no simple relation to the sequence of described occurrences. It seems reasonable, then, to avoid the partial notion of prediction altogether in this connection and suggest the full potentialities of the pattern when *A* and *L* are given, by assigning a temporally neutral term to the derivation of *B*, say 'positing'.³ From assumed laws and information about some

¹ *b* may, it goes without saying, have occurred at any time relative to the givenness of *A* and *L* as well.

² We might call one or both of these inferences 'postdictive', following Reichenbach in *Philosophic Foundations of Quantum Mechanics*, California, 1944, but, contrary to Hempel and Oppenheim, op. cit. p. 138, 'postdiction' is not applicable to explanation.

³ Our use of this term should be clearly distinguished from other uses in the literature, especially that of Reichenbach in his many discussions of probability and confirmation, e.g. in *Experience and Prediction*, Chicago, 1938.

spatio-temporal regions, we posit phenomena at other such regions, in any spatial or temporal relations to our assumed phenomena or our own utterance. Some positing is also predicting, but prediction has no more primacy for the pattern in question than positing events to the left of us in space has. Whether, aside from the pattern, there is independent reason to consider prediction scientifically or epistemologically primary is a question which we shall discuss at a later point.

(b) Now consider explanation, i.e. with *B* given, and a suitable set of statements *A* and *L* provided afterwards, fulfilling Hempel and Oppenheim's four requirements. Once again we find that, since no temporal criteria are to be found among these requirements, they define a wider class of inferences than simply explanatory ones. A non-explanatory instance which fits the pattern is afforded by any case where *b* precedes *a*. A concrete illustration is at hand in any situation analogous to our previous example (ii). Thus, given a description of *some eclipse in former times b*, an astronomer who provides appropriate laws *L* and statements describing *some relevant configuration of celestial bodies during his own lifetime a* from which *B* is deducible, is fitting the pattern, but is surely not *explaining* or *accounting for b*. For explanation, we require, in addition to our four desiderata, that *a* must not temporally follow *b*.¹ What is common to our above instance and explanation is not the temporal order of *a* and *b*, but rather the givenness of *B* and the later provision of *A* and *L*. To suggest the full potentialities of the pattern in such a case, we ought to drop the partial notion

¹ This does not imply that all these requirements are sufficient even if necessary. For example, suppose symptom *S* precedes and is a lawfully sufficient condition for contraction of cancer, enabling prediction. Yet we do not, it might plausibly be said, *explain* contraction of cancer by the presence of *S*. It may be noted that, while we require that *a* not follow *b*, we have not also required that *b* precede the explanation-utterance accounting for it. Since Hempel and Oppenheim interpret the givenness of *B* as implying that *b* precedes its explanation-utterance, it may be worthwhile to justify here our departure from this view: The givenness of *B*, it seems to us, means in practice, merely that we are fairly confident in its truth, but such confidence is surely not limited to statements about the past. If common use is any guide here, then if asked 'Why will the sun rise tomorrow?' I may reasonably be said both to be predicting and explaining the sun's rising when I offer the appropriate astronomical information. Of course, I cannot here be certain of the *truth* of *B*, which truth is necessary if I am truly to explain. But the same uncertainty holds for a *B* which refers to the past, though I am confident in its truth. It seems then that there is here no sharp temporal difference crucial to explanation, though it is required that *b* must not precede *a*, whenever *b* occurs.

EXPLANATION, PREDICTION, ABSTRACTION

explanation' here and again assign a temporally neutral term to the provision of *A* and *L*, e.g. 'substantiating'.¹

(c) If both explanation and prediction are characterised by temporal asymmetries which are irrelevant to the generality of scientific inference, is there any independent epistemological ground for considering explanation or prediction as central to scientific procedure?

(i) It may be noted that the interpretation of explanation in question is generally taken as a reflection of causal notions, and that the peculiar temporal asymmetry of explanation is identical with the temporal asymmetry of cause and effect. To consider the latter notions central to science is to justify treating explanation as scientifically primary.

¹The distinction between explanation and our case cannot be easily discerned by attending to the ordinary use of 'Why...?' as a clue. For while explanations often answer the question 'Why *B*?' in the sense, 'In accordance with what laws and following what conditions does *b* occur?', our case answers the question 'Why *B*?' often asked in the sense, 'What rational grounds are there for asserting '*B*'?'. Perhaps the confusion of these two senses is at least partially responsible for the notion that 'functional' or 'teleological' explanations are explanatory in the ordinary sense, whereas they actually share the logical form of our non-explanatory instance. For Nagel, in 'Teleological Explanation and Teleological Systems' in *Vision and Action*, ed. S. Ratner, 1953, for example, the statement, 'The function of chlorophyll in plants is to enable plants to perform photo-synthesis' is equivalent to 'A necessary condition for the occurrence of photosynthesis in plants is the presence of chlorophyll'. In his words, 'A teleological explanation states the *consequences* for a given biological system of one of the latter's constituent parts or processes; the equivalent non-teleological explanation states some of the *conditions* under which the system persists in its characteristic organization and activities'.

Now it is clear that in Nagel's use here, 'necessary condition' means 'necessary non-subsequent condition' and 'consequence' is temporal in reference. Otherwise, knowing that all breathing organisms die, we might say that since death is a necessary condition for breathing in organisms, the function of death is to enable organisms to breathe. Hence, for his view presumably, to explain *b* functionally or teleologically is to refer to a suitable *later a* such that *A* and *L* imply *B*. Actually, his equivalent non-teleological statement is simply *L* and does not alone provide deductive grounds for asserting the presence of chlorophyll, and hence for explaining it. But assuming that we add the appropriate *A*, i.e. 'Photosynthesis occurs at *t* in plant *P*', this does not *explain* why chlorophyll occurs at some earlier time in *P*, though it *substantiates* such occurrence. We ought therefore not to speak, as Nagel does here, of 'equivalent non-teleological explanations'. Given a case of photosynthesis at *t* however, we may ask 'Why must chlorophyll have been present?', though we do not intend to ask, 'Following what antecedent conditions and in accordance with what laws did chlorophyll occur?'. Confusion of the two questions may partly account for the misleading idea of functional *explanation* as an answer to some 'Why...?'

A number of authors have however remarked the fact that causal notions come to be used less and less by an advancing science, while they remain of relatively constant importance in practical affairs. The point often made in this connection is to stress the relation of causal notions to interest in control by voluntary action. As Braithwaite has recently expressed it,¹

If an earlier event's occurring is a nomically sufficient condition for a later event to occur, we can (in suitable cases) ensure that the later event should occur by taking steps to see that the earlier event does occur. For this purpose it is irrelevant whether or not the later event's occurring is a nomically sufficient condition for the earlier event to occur. . . . But, if a later event's occurring is a nomically sufficient condition for the earlier event to occur, we cannot indirectly produce the earlier event by producing the later event, since by the time that we should be producing the later event the earlier event would irrevocably either have occurred or not have occurred. This difference between the case of regular sequence and that of regular precedence is, I think, the reason why we are prepared to call a nomically sufficient condition for an event a cause of that event if it precedes the event but are not prepared to call it a cause if it succeeds the event.

If this general account is true of causal notions, it would seem to apply equally to so-called causal explanation, which is characterised by the same temporal asymmetry. If scientific inference, however, unlike voluntary control of the future, may be based on temporally backward as well as forward nomic regularities,² it is misleading and partial to view science from the vantage point of such control, and of its cognate notions, 'cause' and 'explanation'. It would seem a better reflection of the full generality of scientific reasoning if we view it as concerned with comprehensive nomological relations among events and abstract from causal explanation entirely. Science may then be compared, in Toulmin's apt analogy,³ to a route-neutral map, quite general as regards direction, but capable of guiding variant itineraries for those with practical purposes.⁴

¹ R. B. Braithwaite, *Scientific Explanation*, Cambridge, 1953, p. 313

² History, geology, archaeology, astronomy are only some striking instances of scientific use of backward regularities. Backward inference, whether postdictive positing or the substantiation (by use of information about the present) of past events, is a partial goal of all sciences, it seems to me, including physics.

³ S. Toulmin, *The Philosophy of Science*, London, 1953, p. 121

⁴ In addition, the connection of explanation with control may indicate why, as pointed out above, temporal requirements alone may not be sufficient to explicate it.

EXPLANATION, PREDICTION, ABSTRACTION

(ii) Much of what has been said of the relation of explanation to voluntary control holds in an obvious way for prediction. In addition, however, the scientific primacy of prediction is often supported by reference to the acceptance or confirmation of statements.

It may, for instance, be granted that we posit events both past and future to our posit-utterances but it is pointed out, for any present posit-utterance, it is peculiarly contingent on the future since there is a possibility that it is reasonably rejected then, owing to future rejection of some confirmatory sentence asserting a future occurrence. But, in the first place, such future rejection may be due to future rejection of some confirmatory sentence asserting an occurrence prior to the present posit-utterance. And, in the second place, it is equally true for any present posit-utterance that there is a possibility of its reasonable rejection in the past. That something is a posit-utterance at t implies neither its acceptance at all times following t nor its acceptance at all times preceding t .

It may then occasionally be suggested that the predictiveness of any given posit-utterance is not a matter of its own future acceptance but involves rather the fact that it is false if any confirmatory sentence asserting a future occurrence is false. Obviously, however, it is also false if any confirmatory sentence asserting a prior occurrence is false. It may, of course, be held that a posit-utterance, all of whose confirmatory sentences asserting prior occurrences are true, is false only if some confirmatory sentence asserting a future occurrence is false. But the obvious converse is equally the case.¹

It may, however, finally be countered that we cannot now voluntarily choose to carry through a past test, while we can now decide to institute a test of a specified posit-utterance in the future. This claim, true enough, is as trivial as the general truth that voluntary control at a given time is of later phenomena, never of earlier; we cannot now choose to institute any past event, *a fortiori* we cannot now choose to institute a past test. There is no special relevance to science in this truism. Furthermore, we have seen that a given posit-utterance is false if a confirmatory sentence asserting a prior occurrence is false, even if all confirmatory sentences asserting future events are true. Hence, even if all its tests which are voluntarily choosable (hence in the future) at t are positive, i.e. yield true confirmatory sentences asserting later events, the posit-utterance at t may still be false.

¹ I am throughout this passage following the usual type of argument in talking of confirmatory sentences, etc., but this does not commit me to such an epistemology.

Pragmatists and positivists have championed the further doctrine that the meaning or content of a physical-object statement *is* its future verifiability. Hence, e.g. even an apparently retrodictive posit-utterance of historian *H* in 1950, 'Caesar crosses the Rubicon' is really *about* future possible confirmations or disconfirmations in experience. This doctrine of meaning is however, ambiguous, since 'future' is unclear.¹ In our example above, are the future confirmations future to 1950 or future to Caesar's crossing the Rubicon? Only if they are future to 1950 is it plausible to construe the content of *H*'s assertion as its testing future to the assertion, but this interpretation leads to quite undesirable results, e.g. a replica of *H*'s utterance in 1954 has a quite different content. If, on the other hand, 'future' here means 'future to Caesar's crossing the Rubicon', then it refers also to confirmations prior to *H*'s utterance, and the ground for considering its predictive content primary disappears. It goes almost without saying that once we are prepared to admit confirmations prior to *H*'s utterance, there is no longer any advantage in excluding those which precede the historical event itself.

But the general difficulties of this dictum on meaning far outweigh, in importance the ambiguity mentioned.² Abandoning it in favour of some other criterion of meaning, we remove a reason for considering prediction as primary which has been dominant in recent philosophy.

3 *The Claim of Idealism*

It is commonly said that scientific inquiry abstracts from its raw subject-matter, and more seems often intended by such statement than the harmless truths that science is embodied in language, that it deals with only some of the concrete entities available for study at any given time, and that it selectively analyses complexes into parts. Often the point seems rather to be that the true *objects of*, e.g., scientific explanation³ are not concrete events or things at all, but supposed

¹ For a discussion of this point, in relation to C. I. Lewis's analysis of historical statements, see the writer's 'Verifiability in History: A Reply to Miss Masi', *Journal of Philosophy*, 1950, 47, 164 ff.

² For a critical review of these difficulties leading to an alternative proposal in terms of translatability into some empirical language, see C. G. Hempel, 'Problems and Changes in the Empiricist Criterion of Meaning' *Revue internationale de philosophie*, 1950, II, 41-63.

³ For clarity, I here discuss explanation specifically, but the transfer of this treatment to prediction is readily made.

abstract, intensional features, or properties of the latter, idealistic entities like phenomena, facts, or states-of-affairs.

Now this view might plausibly be construed as claiming that every adequate analysis of locutions like ' . . . scientifically explains—', or ' . . . scientifically accounts for—' presupposes an abstract, intensional ontology. Independently of whether the sciences themselves afford evidence of the existence of idealistic entities, that is, the *theory* of scientific explanation is asserted to require them.

It will be recalled that in our previous discussions of the supposed inference-pattern of scientific explanation, we followed the conventional description of it, i.e. we spoke of accounting for or explaining the *event* *b* by providing appropriate sentences *A* and *L* which have its description *B* as a logical consequence. If the idealism claim is correct, however, all talk of this pattern as concerned with explaining or predicting the concrete events mentioned in the explanandum is, taken literally, false (unless it is merely abbreviatory of appropriate idealistic talk). In what follows, we shall (a) discuss grounds for the idealism claim as formulated above (b) propose an inscriptional alternative, and (c) specify how this alternative enables us to speak of science as abstractive, though rejecting abstract entities.

(a) Consider that events (taken in the ordinary sense as concrete occurrences with spatio-temporal boundaries) are describable in alternative, logically independent ways and hence that it is false to suggest that *B* is a *unique* description of *b*. With a specified stock of individual predicates ' *Q* ', ' *R* ', ' *S* ', etc., we could truly describe *b* by either affirming or denying for *each* of these predicates that it is true of *b*. If so, how can we speak of having explained the *event* *b* when we have fulfilled our pattern's requirements in providing appropriate deductive grounds for just one of these true descriptions, say ' *Rb* ' ? Have we both explained and not explained *b* in providing such grounds for ' *Rb* ' but not for the equally true description ' *Qb* ' ? Clearly, a particular fulfilment of the pattern explains not simply the event *b*, but *b* as qualified in a certain specified way, or the fact that *b* is so qualified.

Suppose that, regarding such an event as Cicero's birth (in 106 B.C. on the particular day *D*), we provide specific information to the effect that the relevant conception occurred 280 days prior to *D*, that it was followed by a normal pregnancy, was human, etc., as well as the generalisation that, for human beings, births follow their respective conceptions, in normal cases, by exactly 280 days. Have we now explained Cicero's birth ? But this birth could be characterised in innumerable

ways other than in terms of when it occurred, e.g. it followed a specific period of labour, it had individual obstetrical characteristics, etc. Clearly our particular explanans does not explain Cicero's birth as such, but rather Cicero's-birth-as-occurring-on-*D*-in-the-year-106 B.C., or as often expressed, the fact that Cicero was born on *D* in 106 B.C. We ought then to discard simple contexts like 'Explanans *E* explains', (where the dots are replaceable by names of concrete occurrences) in favour of indirect locutions calling for appropriate qualification of the occurrences in question.

For our above case (calling its specific explanans '*E*''), we might say :

(i) *E*' explains why Cicero's birth occurs on *D* in 106 B.C.

Now in view of the fact that :

(ii) Cicero's birth = Tully's birth

is true, whatever we say of Cicero's birth ought to be truly sayable of Tully's birth. However, when we replace 'Cicero's birth' by 'Tully's birth' in (i), we turn its truth into falsehood, since though *E*' logically implies 'Cicero's birth occurs on *D* in 106 B.C.', it does not logically imply 'Tully's birth occurs on *D* in 106 B.C.' Following Quine,¹ we may say that the occurrence of 'Cicero's birth' in (i) is not purely referential, i.e. the truth of (i) depends not merely on the particular event, but on the way in which the event is denoted.

Since, moreover, the identical (as well as any appropriately analogous) inter-change fails to alter the truth of the subsentence of (i), i.e. :

(iii) Cicero's birth occurs on *D* in 106 B.C.

when it stands alone, but does alter truth when it is embedded in (i), we may, again following Quine,² call the contexts exemplified by (i) referentially opaque, to signify that, like quotation contexts, they can change referential occurrences into non-purely-referential occurrences.

The important thing to note about referentially opaque contexts is that we cannot quantify *into* them from the outside, just as we obviously cannot do so for quotation contexts. Thus to apply existential generalisation to the occurrence of 'Cicero's birth' in (i), we should get :

¹ W. V. Quine, *From a Logical Point of View*, Cambridge (Mass.), 1953, p. 140

² Ibid. p. 142, and W. V. Quine, 'Three Grades of Modal Involvement', *Proceedings of the XIth International Congress of Philosophy*, 1953, 14, 67

EXPLANATION, PREDICTION, ABSTRACTION

(iv) $(\exists x) (E' \text{ explains why } x \text{ occurs on } D \text{ in } 106 \text{ B.C.})$

or :

(v) Something is such that E' explains why it occurs on D in 106 B.C.

But what event is it whose time of occurrence is explained by E' ? Cicero's birth, or equivalently, Tully's birth? But we have already seen that to put 'Tully's birth' into (i) makes it false. The failure of (iv) and (v) means thus that we cannot construe sentences like (i) as being *about* explained events; we cannot, that is, analyse them in quantificational form so that the bound variables take the events mentioned in the explananda as values. Correspondingly, we cannot then describe our pattern as concerned with explaining these concrete events at all.

We may thus be inclined to construe the objects of scientific explanation as so-called facts or states-of-affairs, which cannot be denoted by non-synonymous units in the nature of the case, and hence do not give rise to the difficulty of referential opacity as elaborated above. We may, for example, construe the whole subsentence as it occurs in (i) as designating an abstract fact or state-of-affairs (Cicero's birth occurring on D in 106 B.C.) not equally designated by 'Tully's birth occurs on D in 106 B.C.' in the same context, and existentially generalise as follows :

(vi) $(\exists x) (E' \text{ explains why } x)$

or :

(vii) Something is such that it is explained by E' (i.e. the state-of-affairs Cicero's birth occurring on D in 106 B.C.).¹

(b) The referential opacity of contexts represented by (i) precludes analysing them as referring to concrete entities mentioned in their respective explananda. In this sense, then, we cannot take these concrete events, processes, or other things to be the objects of scientific

¹ In 'The Function of General Laws in History', Hempel seems indeed to be taking some such view. He there construes events or phenomena not as concrete minimal spatio-temporal chunks, but rather as 'kinds or properties of events', while descriptive sentences like A and B figuring in explanantia and explananda express the *fact that* such properties occur at specific places and times. One consequence of this particular analysis seems to be a revision of the ordinary use of 'event' to denote properties of what are usually called 'events'. Insofar as the distinction between the two may be blurred, this consequence may be thought undesirable. Also, it would seem that 'event' in this revised use is not appropriate for denoting causes or effects, a goal which some proponents of the explanatory pattern discussed seem to have had at the back of their minds.

explanation. If we persist in this sort of interpretation, we prepare the ground for those familiar and puzzling philosophies of science according to which science cannot ever explain its own true object of study.

Nevertheless, an idealistic interpretation such as is given in the paragraph before the last is not the only alternative open to us. We do indeed have to avoid quantifying into referentially opaque contexts in the manner of (iv) and (v), but we can, anyhow, suggest an analysis of (i) which remains concrete in ontology throughout, though referring to linguistic rather than extra-linguistic entities. This analysis rests on the fact that what the explanation-pattern under discussion actually exhibits is a logical relationship among certain statements, i.e. explanantia and explananda. Assume, then, that the whole why-clause in (i), i.e. 'why Cicero's birth occurs on *D* in 106 B.C.' functions as a single indivisible predicate in this context, applicable *not* to any so-called state-of-affairs, but rather to every concrete inscription which is a rephrasal or translation of the subsentence following the 'why'.¹ Now analyse (i) as :

- (viii) $(\exists x)$ (x is a why-Cicero's-birth-occurs-on-*D*-in-106 B.C., and x is logically implied by E' which fulfils (R_2) , (R_3) , and (R_4)).

The variable here ranges over concrete inscriptions exclusively, and ' E' ' can be taken to name some such inscription as well (with the relation of logical implication appropriately understood). Construing the why-clause as applying to *rephrasals* rather than simply to replicas of 'Cicero's birth occurs on *D* in 106 B.C.' (like direct quotes), enables (viii) to cover cases where the explanans is formulated in some language other than that of (i).

As we focus our attention on some particular branch of science, moreover, taken as formulated in some specific constructed language, and where our purpose is to build a theory of explanation for this language rather than to explicate ordinary statements like (i), it becomes possible to avoid interlinguistic reference (as in (viii)), together with such rather imprecise notions as those of translation or rephrasal.² We may

¹ This interpretation was proposed originally for indirect discourse generally in the writer's 'An Inscriptional Approach to Indirect Quotation', *Analysis*, 1954, 14, 83-90.

² For a discussion of some of the difficulties in using synonymy and translation to explicate indirect discourse, see the writer's 'On Synonymy and Indirect Discourse', *Philosophy of Science*, 1955, 22, 39-44.

EXPLANATION, PREDICTION, ABSTRACTION

then always specify particular inscriptions in the language in question, and interpret explanation in terms of the relation of logical implication between some such inscriptions. In effect, we may look upon both (viii) and the latter approach as the assimilation of event-explanation to the explanation of laws or generalisations, mentioned earlier in section (I). For the explanation of laws is clearly no more than a question of relating law-statements to appropriate other statements.

(c) This assimilation of event-explanation to that of laws or generalisations serves to underline the fact that *abstractiveness* is distinct from and need not involve *admission of abstract entities*, i.e. for a mode of inquiry to be said to abstract from raw experience or from the world, it need not be thought to take abstract entities as its objects. Clearly, the explanation of laws or generalisations is no less abstractive relative to experience than event-explanation, though conceivable in straightforward fashion as taking concrete law-sentences for its objects and concerned with showing appropriate relationships among certain of these objects. Its abstractiveness relative to entities referred to by its explananda consists just in the multiplicity of explananda with the same reference (i.e. correlated by reference with just the same entities) and the consequent need of *selection* in any case of explanation. If every law, that is, makes the same reference to everything via the universal quantifier, each law-explanation abstracts from everything in selecting as explanandum only one out of many sentences, all equally referring to everything. Such selection in no way depends on these initially choosable sentences being abstract.

Analogously, though event-explanation be taken (in accordance with (viii) and the paragraph following) as not requiring abstract entities for its objects, it is yet thoroughly abstractive relative to the events mentioned in its explananda. For each such event is correlated with many sentences, i.e. all those expressing true descriptions of it. In the selection of one of these as explanandum on a particular occasion lies the abstractiveness of event-explanation. As before, the crux of the matter is the correlation of a multiplicity of choosable units with the same entities rather than the abstractness of these units.

The University
Harvard

DETERMINISM AND PREDICTABILITY *

D. J. O'CONNOR

It has often been said that the concept of determinism can best be explained by agreeing to call an event determined if it is predictable in principle. The notion of predictability has been used in this way to elucidate the supposedly vaguer notion of determinism by philosophers of such diverse views as Karl Pearson, Eddington, Planck, Schlick, and Popper. Schlick, for example, explains the notion as follows. To say that *A* determines *B* means no more than 'given *A*, *B* can be computed'. More explicitly, 'there is a general formula that yields a description of *B* as soon as certain values from the initial conditions *A* are inserted in the formula and certain variables in the formula, for example, the time *t*, are given a definite value'.¹ If such views are right, it is reasonable to try to explain the meaning of 'determinism' by explaining the phrase 'predictable in principle'. Even if they are wrong, as I think they are, it is still a useful starting point.

What do we mean when we say that events of a certain type are predictable? The standard examples of such events are those which fall within the scope of a well-developed science like physics or astronomy. It is common knowledge that the range of events which can in practice be predicted is very limited. Nevertheless, it has sometimes been suggested that these limitations can be removed by increasing our knowledge of the laws of nature or by improving the techniques of observation and measurement by which we collect our data or by perfecting our methods of calculation. It is supposed, in other words, that these limits are mere *de facto* practical restrictions on our present knowledge and are not restrictions 'in principle'. If we ask what is involved in this claim and, in particular, what is the meaning of the phrase 'in principle', it is very difficult to get a clear answer. I shall therefore start with the much easier question: What conditions must be satisfied before we should accept a claim that a certain event had in fact been predicted? It is simple enough to list

* Read to the Philosophy of Science Group, 14th February 1955

¹ M. Schlick, *Gesammelte Aufsätze*, p. 73

DETERMINISM AND PREDICTABILITY

the essential conditions. To justify a claim that I had predicted a certain event E which had occurred at time T_1 I should have to show (a) that at time T_0 , prior to T_1 , I had described E and stated that it would occur at T_1 ; (b) that at time T_0 I possessed evidence which justified the prediction and (c) that I had actually deduced from this evidence, by a valid process of inference, the statement embodying the prediction.

Clearly, if only (a) was satisfied, we should not be willing to say that E had been predicted in any scientific sense. We should rather call such a forecast 'prophecy', 'precognition' or 'clairvoyance', if indeed we did not just suppose it to be a lucky hunch. If both (a) and (b) were satisfied, we should still want to know that the statement embodying the prediction had been drawn from the evidence by a legitimate process of reasoning. Thus predictions may vary in a number of ways. They may be more or less accurate in their description of the event and in the time to which it is assigned; and the description may be more or less complete. Moreover, predictions may be the outcome of varying sorts of inference. A critic might be prepared to deny to any ostensible prediction the status claimed for it on the ground that the description of E or the time to which it was assigned was too vague or too inaccurate or because the argument by which the prediction was drawn from its evidence was logically unsatisfactory.

These criteria are reasonable enough but they are quite imprecise. Moreover, as I shall try to show, their imprecision is such that we have no way of removing it. As a consequence of this, we cannot maintain the close connection which is commonly supposed to exist between the notions of determinism and predictability, though we do, of course, settle on pragmatic grounds the criteria which are acceptable in practice in any branch of science. Any ostensible prediction may satisfy to a greater or less degree the standards I have mentioned. There are no well-established rules of logic or of common use which can enable us to say whether a given candidate to the title is a genuine prediction or not. We would like to have rules that would give us a definite *yes* or *no* to the question: Was it a genuine prediction (and not, for example, a wild guess or a shrewd one or even an instance of clairvoyance or precognition)? But such a decisive test is not available to us and never could be. Predictions are not natural features of the world which we recognise and classify. They are human performances and it is we who decide what shall count as such

a performance. In a somewhat similar way, we can recognise birds as partridges, rooks, or magpies, but have ourselves to decide what shall count as game birds or as vermin. Such decisions are not arbitrary, but they are decisions.

So far I have tried to state the conditions which have to be satisfied before we can say that a genuine prediction has been made. I want now to look at the more elusive notion of 'predictability in principle' and its supposed relevance to determinism. We are faced at once with the difficulty that, as I have tried to show, the general concept of predictability is a vague one simply because it is not certain where we are to draw the boundary between those statements which are to count as predictions and those which fail to qualify. Yet the notion is derived from the fairly precise one sketched by Laplace in his well known passage about the Supreme Calculator. If we have an isolated physical system of a finite number of particles and we know their respective positions and momenta at any given time together with the laws of mechanics, we can deduce any future state of the system from this information and indeed any past state as well (The time series is supposed to be symmetrical in this respect, even though we do not talk of past states of the system being 'determined' by its present state.) We may waive for the moment the considerable difficulties hidden in the phrase 'we can deduce any future state . . .' It is plain at any rate that all forms of causal connection do not conform to this pattern simply because, in many cases, we have no means of expressing the relation between antecedent and consequent states of affairs as a functional relationship of the type usual in mechanics. If, for example, we forecast the distance a freely falling body will have travelled at a certain instant by making an appropriate substitution in the formula $s = \frac{1}{2}gt^2$, we are at least in no difficulty over the type of reasoning which leads to our prediction. The rules are the rules of mathematics. But even if we take examples of this sort of prediction, which are the cases most favourable to the equation of determinism and predictability, it is easy to show that the concepts cannot be equivalent.

We can show this by considering the nature of measurement. Measurements are descriptions in terms of quantity. They vary in accuracy and we do not suppose that measurement of any continuous quantity can be accurate in any absolute sense. We assume that there are limits of error appropriate to any given measurement and often these limits are stated. The bearing of this well-known fact on the

DETERMINISM AND PREDICTABILITY

notion of predictability is that no predictive assignment of measurements to an event can identify an event uniquely. Although probably only one event will occur in the spatio-temporal region assigned by the forecast, a very large number of other events *might have satisfied* the prediction. They would have done so if their dimensions fell within the threshold of accuracy embodied in the measurements.

If this is so, it follows that to say that an event is predictable is not so strong a claim as to say that it is determined. For no prediction can distinguish the predicted event in such a way as to discriminate it from any of the other possible events that could fall under the same set of measurements. To take a simple example, a measurement of 10 centimetres plus or minus 0.001 centimetres would be satisfied by any length falling within the range 9.999 to 10.001 centimetres. There can be an indefinite number of such lengths though we cannot distinguish them without finer limits of error. Thus where we predict measurable features of our experience, our predictions can do no more than specify a *class* of possible events. If one of these events is observed to occur, the prediction is counted as successful.

I have considered the sort of case most favourable to the claim that all events in the universe are 'predictable in principle'. It is easy to see that even here determinism and predictability are not terms which are co-extensive in their range of application. The connection between them becomes still more tenuous when we take account of events outside the province of physics. Once we formulate our predictions in non-metrical terms, as we must do for most events, we are in difficulties over the sort of inference that leads to the prediction. Moreover, we meet, in an aggravated form, the same difficulty that we find in the case of measurement; the statement which expresses our prediction is never capable of identifying without ambiguity one and only one event whose occurrence would satisfy the prediction. For a description can do no more than specify a *class* of closely similar events, whose differences lie beneath the threshold of discrimination which the detail of the description provides. We can indeed make this class smaller and smaller without limit by making our description more and more detailed. But however far we go, it is a necessary consequence of the nature of language that we can never make the description perfectly determinate. For it is an essential feature of a descriptive word or phrase that it cannot *identify* anything. If it did, it could not also describe. And it is an empirical fact that we can never make our descriptions complete in every detail. However detailed

and delicate a description may be, it is only a variable that can be satisfied by any of a certain range of states of affairs.¹

I can sum up this discussion by saying that while every event must be perfectly determinate, no prediction can identify such a determinate event unambiguously. Thus if we really believe that predictability and determinism are equivalent notions, we shall have to admit that determinism is a much weaker concept than it has usually been taken to be. There would then be a loose-jointedness in nature so that a given state of the world at an instant determined not a given subsequent state but a class of possible states ; and there would be no explanation why one rather than another of these states should actually be realised. But to make a supposition of this sort merely because there is a limit to the precision with which we can formulate our predictions would clearly be unjustified.

I do not wish to deny that it is legitimate to make a methodological equation of the two concepts for the purpose of refuting determinism. Popper does this in his well-known papers on indeterminism in quantum and classical physics.² By showing that universal predictability is an unattainable ideal, he refuted the suggestion that every event is determined. But this refutation holds only against those who believe that predictability and determinism are either interchangeable concepts or at least so closely related that we know that determinism entails predictability. It seems to me, however, that once we realise the basic differences between the two concepts, it is no longer plausible to maintain that we can explain one in terms of the other. To use rather old-fashioned language, predictability is an *epistemological* concept which takes its meaning from the operations involved in the collection and assessment of evidence. Determinism is an *ontological* concept which purports to define the extent to which events are causally inter-related. While determinism entails a one-one correspondence between antecedent and consequent states of affairs, predictability entails a one-many relation between the prediction and the possible events which would satisfy it.

To call certain kinds of event 'predictable in principle' instead of simply calling them 'predictable' is a way of blurring an important distinction. Prediction is essentially a practical procedure of making

¹ It has been pointed out to me by Dr H. R. Paneth that this argument would fail in its application to spatio-temporal measurements if we were to suppose a quantised space-time. But I do not know of any evidence in favour of such a hypothesis.

² K. R. Popper, this *Journal*, 1950, I

DETERMINISM AND PREDICTABILITY

observations and deductions from observations in accordance with established rules. The legitimacy of these techniques of deduction and observation is justified in the event by the success of the prediction. Once we have settled what is to count as a prediction, the question whether a future event is predictable is a question of fact to be decided by ordinary inductive reasoning. Knowing what events have been predicted, we can state with varying degrees of assurance what other events will probably fall into the same class. To say that a certain event, which we cannot in fact predict, is nevertheless 'predictable in principle' is merely to make the hypothetical statement: if we had the relevant evidence and could make suitable deductions from it, we should be able to describe the event in advance and say when it would occur. But this hypothetical does not even look informative. There is no important difference between the meanings of 'predictable in principle' and 'predictable in practice'. The first class of events is larger than the second only to the extent that reasonably foreseeable improvements in our predictive techniques can bring events which cannot at present be predicted within our range.¹

University of Liverpool

¹ It may be worthwhile mentioning an ambiguity in the English word 'determine' (and in some of its foreign equivalents) which may have encouraged the mistaken identification of predictability and determinism. If we predict an event, we determine it in the sense that we identify it as a result of inference. It is in this sense that astronomers determine the dates of eclipses or coroners' inquests determine causes of death. But, in addition, the word can mean 'cause'. And though we can determine a disputed question of fact, we cannot in the same sense determine the fact itself. This ambiguity makes it easy for us to suppose that if an event can be determined in the cognitive sense of the word, it must also be determined in the causal sense. This is sometimes true but not so often that we can identify the two concepts or use one to explain the other.

SOME ASPECTS OF PROBABILITY AND INDUCTION (II)*

JONATHAN BENNETT

3 *Eliminative Induction*

In dealing with various accounts of induction, other than his own, Kneale is led to criticise very thoroughly Keynes's views on induction as a process of increasing the probability of a given hypothesis by the elimination of possible alternatives. His criticisms fall into two groups : one consisting in a pair of arguments against Keynes's theory in particular, and one consisting in a pair against any attempt to explain the confidence that we feel in inductive results as being based on the theory of chances. There is more to be said on each of these points, and they will be taken in order.

(i) Keynes's theory of eliminative induction is too well known to need more than the sketchiest outline in this paper. Briefly : given several occurrences of situations each characterised by the qualities $a_1, a_2, a_3, \dots, a_n$, the question may arise as to whether some of these qualities, say a_1, a_2, a_3 , are causally bound up with some of the others, say a_{n-1}, a_n , in such a way that wherever the former group of qualities are found the latter will be found also. What we do in such a case, says Keynes, is to apply the method of difference, finding as many cases as possible of a_1, a_2, a_3 , which differ from one another as much as possible in respect of all the qualities a_4, a_5, \dots, a_{n-2} . If one or more of these cases lacks one or both of the qualities a_{n-1}, a_n , then our question is answered in the negative. But so long as this does not happen, the more such cases we can find the more alternative possibilities we eliminate—possibilities, that is, such as that a_1, a_2, a_3 produce a_{n-1}, a_n only in the presence of a_4, a_5 ; or that the question of a thing's being a_1, a_2, a_3 is entirely irrelevant to the question as to whether it is a_{n-1}, a_n , these latter qualities having been caused in the original situations by a_7, a_8, a_9 . The more such alternatives we eliminate, the greater is the probability that our original hypothesis is true.

* Part I of this paper appeared in the November 1956 Number

SOME ASPECTS OF PROBABILITY AND INDUCTION

The first criticism which Kneale makes of this doctrine is somewhat obscure, but it appears to be as follows. If the characters α and β are found to be frequently conjoined and if we wish to discover whether there is a causal relation such that all α 's are β 's, then, on the theory, we must attempt to multiply instances of α -ness which are β in such a way that finally our instances have nothing in common except that they are all α and β ; and then we can assert that all α 's are β 's, or that α -ness necessitates β -ness, or something of the kind. 'But', says Kneale, 'we should presumably wish to reach exactly the same position if we were trying to establish by this method that all β things are α ' (p. 209). This, he thinks, shows that at best induction by elimination can serve only to establish reciprocal connections, and since Keynes 'does not think of all scientific generalisations as simply convertible' this constitutes a refutation, if not of his account of the method, at least of the place which he gives it in his total theory of scientific procedure.

But all this is simply a mistake. If we wish to establish that α -ness causes β -ness, we try to vary as much as possible the *instances of α -ness* that we collect, in the hope that no variation will result in our finding instances of α -ness which are not instances of β -ness. What *other* instances of β -ness there may be will be a matter of no concern for our enquiry. It is made easier to get these facts wrong by talking in terms of the qualities of things rather than of the qualities of events. When the scientist 'collects instances' he experiments—he *makes* instances of α -ness and then investigates them to see whether they are also instances of β -ness; and this is a quite different activity from creating β situations and investigating them for α -ness, as any experimental geneticist knows.

Kneale is led into this unjust criticism of eliminative induction not just by the fact that, apparently, he puts too much weight on Keynes's use of the symmetrical concept of 'analogy' and consequently overlooks the fact that experiment has *direction*, but also by what he seems to intend as a basic, theoretical argument to show that the eliminative method is at best restricted to the discovery of reciprocal connections: 'If we try to show that α -ness necessitates β -ness by eliminating other characters which might at first be supposed to necessitate β -ness but are not in fact found in all our instances of β -ness, we need two premisses. First we must assume that some character found together with β -ness in our instances necessitates it. And secondly we must assume that nothing which is absent in any instances where β -ness is

present can be a character necessitating it. But this second assumption is just the doctrine that all necessitation is reciprocal' (p. 209).

We may grant that the first premiss, or something like it, is required; but what of the second? Why can we not try to establish the law 'If α , then β ' (to use a convenient short-hand) by eliminative induction, at the same time granting that 'If γ , then β ' may also be true? In such an eventuality, it might be that none of our cases of α -ness are cases of γ -ness at all (α and γ might even be incompatible characters), in which case the induction would not be concerned to eliminate γ -ness as a possible cause of β -ness: it would be concerned only to eliminate possible causes which are 'possible' in the sense that they *do* occur in some of our $\alpha\beta$ cases. And if some instances of α -ness (and β -ness) also have the characteristic γ , the induction can still proceed in an attempt to establish that α -ness is a *sufficient* condition of β -ness: even if 'If γ , then β ' is true, we eliminate γ in order to show that in the given cases of $\alpha\beta$ -ness the characteristic γ is irrelevant *to our present purpose of testing* 'If α , then β ', irrelevant in the sense that the α situation would have been β even were it not γ .

To keep the record straight, it should be noted at this point that Kneale's views about causation as involving principles of necessitation in no way imply that we must think of causes as *necessary and sufficient* conditions of their effects. This crude mistake, which is not much better than a pun on 'necessary', has a certain currency even today in idealist circles, but I do not think that Kneale makes it.

(ii) Kneale's other main objection to the eliminative induction theory hinges on the fact that it involves the assumption of 'the limitation of independent variety', the assumption, that is, that in any given case the number of possible alternative hypotheses is finite. Were this not so, the argument goes, then the elimination of half a dozen possibilities in one experiment, and three or four more in the next, would advance us no distance at all in increasing the probability of the hypothesis which we are trying to prove: for the fraction whose numerator is 1 and whose denominator is the number of remaining possibilities would remain the same. All this is pointed out by Keynes himself, who appears to differ from Kneale in this connection mainly in that the latter lays more stress on the magnitude and unprovability of the assumption. It is true that Keynes goes one step further and suggests that the success of inductions based on the principle of the limitation of independent variety constitutes evidence for (increases the probability of) the principle itself; and we may grant,

SOME ASPECTS OF PROBABILITY AND INDUCTION

though a shade less confidently than Kneale asserts, that this mode of argument for the principle is unsound. But more can be said about the principle than this.

In the first place, it is possible to describe a set of conditions under which the method of elimination would work even if the principle were false. For if we decide to investigate the hypothesis 'If α , then β ' when there is an infinity of possible alternative hypotheses to account for the $\alpha\beta$ situations so far observed, it might be that the very first (or some subsequent) experiment will differ from the previous $\alpha\beta$ situations in respect of an infinity of characteristics, in such a way that an infinity of possible hypotheses is eliminated at one blow. In such a case the probability of 'If α , then β ' *might* be genuinely increased and might be in a position to go on being increased by the further elimination of merely finite numbers of fresh hypotheses. We cannot know that this situation often obtains, but by the same token we cannot know that it does not; and the most plausible persuasions to the latter conclusion also tend in the direction of showing the principle of limitation to be itself true, in which case *cadit quaestio*.

Secondly, there is a class of situations in which the infinity of possible alternative hypotheses may take the form of possible variations, within limits, of some numerical function occurring in the hypothesis under investigation. There is, for instance, an infinity of possible values for x between $x = 7.00$ and $x = 7.01$; and the infinity of characteristics in some sets of $\alpha\beta$ situations on the basis of which we assert 'If α , then β ' where β includes a function involving the number 7, might be such that if they are relevant at all to our hypothesis it is only in that they might be responsible for some deviation within the range 7.00—7.01. In such a case we may also claim that the principle of limitation need not be assumed: so long as the possible alternatives occupy a limited range, however minutely it is *theoretically* possible to divide that range, we have for *practical* purposes only a finite number of alternatives.

Apart from these untroublesome cases where each member of the 'infinity of possible hypotheses' differs from the others only in respect of the name of some real number, all the numbers thus named lying in some fairly brief interval, *do* we ever have an infinite number of possible hypotheses? It is arguable that we do not. After all, we are concerned with hypotheses of a not more than manageable degree of length and complexity—with *usable* hypotheses—and there are good reasons for believing that there is never an infinite number of

these which could cover any given range of experiments. Practical science must use the vocabulary at hand, together with such additions as practising scientists can make from time to time; in a finite period of time only a finite number of additions can be made, and every vocabulary at present available to practising scientists is finite in length; therefore practical science can be usefully thought of only as a discipline having at its command a finite vocabulary. From this it follows that practical science must *either* be restricted to a finite number of statements, *or* tolerate an infinity of statements *each* of which is infinite in length. It is evident—as a trivial corollary of the claim about ‘manageability’ above—that the latter cannot be considered an open possibility for science considered as a human activity. And so we are left with the former alternative—which is the limitation of independent variety—not as a proven truth about the universe but as a condition the failure of which entails not just the failure of Keynes’s theory but also the ultimate failure of science itself. Any condition bearing this relation to science is surely entitled to occupy an axiomatic place—though with no claims made about its truth—in a philosophy of science.

(iii) The first of Kneale’s general objections to ‘all attempts to justify induction within the theory of chances’ is summed up in his statement: ‘It is only reasonable to speak of chances where it is also reasonable to speak of equipossible alternatives. But there can be no alternatives to the holding of a necessary connection or to the holding of a probability relation’ (p. 212). This argument depends on Kneale’s view—which would, I think, be regarded today as highly idiosyncratic by all but thorough Thomists—that causal necessity is all of a piece with logical necessity. On this view the contradictory of a true scientific hypothesis may be ‘conceivable’ but it is not (logically) ‘possible’. How then, the argument runs, can we speak of the elimination of *possible* hypotheses?

We shall here pass over Kneale’s identification of the theory that there are physical laws which are not just conjunctions (which theory he claims to be essential to the theory of eliminative induction) with the quite different theory that these laws are all of a logical nature. For even on its own premisses the above argument will not hold. For there is, as Kneale recognises, such a thing as *possibility-on-the-evidence*. And whatever theory of chances one may construct such that ‘we are not justified in [talking of] second-order chances’ (p. 213) based on ‘second-order possibility’ (i.e. possibility on the evidence), it just

is the case that if we start in a state of knowledge in which any one of 100 statable hypotheses might, for all we know, be the explanation of certain phenomena ; and if we move from this position to a state of knowledge in which there are only thirty hypotheses any one of which could, for all we know, be the required explanation ; then we have raised the likelihood of each one of the thirty hypotheses. It will be readily seen, in fact, that what we have here is just a special case of the frequency theory of probability, and that Kneale is committed to a rejection of it by his private views on the nature of physical necessity together with his special theory of probability—a theory which has here been argued to be untenable. Keynes states the point perfectly : ‘ An inductive argument affirms, not that a certain matter of fact is so, but that *relative to certain evidence* there is a probability in its favour ’ (*Treatise on Probability*, p. 221, italics in original). Nor is it true, as Kneale affirms, that ‘ the project of dealing with induction in this way has been encouraged by the indifference theory, which makes ignorance a sufficient ground for assertions of probability ’ (p. 213). The probability of a statement is, of course, relative to some other statements : statements do not just *have* probability. We should be foolish indeed if we did not relate our assertions of probability to what we know, and fortunate indeed if our state of knowledge were not also a state of ignorance ; but this is a far cry from admitting that eliminative induction is based on anything so crude as the indifference theory.

(iv) It is just possible, however, that Kneale’s reference to the indifference theory is directed not at the whole concept of probability as it is employed in any description of eliminative induction, but against a particular, ‘ fantastic ’ assumption which he says the theory must make : ‘ When . . . we try to work out what is involved in talking of the chances of there being a certain probability relation between two characters, we must first think of the characters as constituting an ordered pair and then suppose that there is some initial probability of this dyad’s exemplifying a certain probability relationship *simply because it is a dyad of characters* ’ (p. 213, italics in original).

Now, this is very curious. In the first place, it is not at all clear why it is that we must talk about the probability that there is a *probability* relationship between, say, α and β . What concerns us, surely, is the probability that there is a *causal* relationship between them. We do not, furthermore, assert that there is a probability of this relationship’s holding between α and β just because they constitute a dyad of

characters. Our grounds are rather that $\alpha\beta$ is a dyad of characters which occurs in one or more natural phenomena, that each term in the dyad has a causal relationship with some character of each phenomenon characterised by $\alpha\beta$, and that it is assumed that there is only a finite number of such characters to be considered. This, far from being fantastic, is a simple example of the kind of use to which we daily put the concept of probability.

'No one', says Kneale, 'believes seriously that the probability of a scientific generalisation or theory could be properly represented by a fraction' (p. 214). On the contrary: there is no reason why the probability, *relative to given evidence*, of a scientific hypothesis should not be expressed in a fraction. This seems impossible not because of some impropriety in the whole idea but simply because we never know what *value* the fraction has. Of course, it would be neither of use nor of interest to us if we did know: we do not need to know what a theory's probability is; we need to know how to increase it.

University of Cambridge

NOTES AND COMMENTS

Criticism of Fairbairn's Generalisation about Object-Relations

FAIRBAIRN puts forward the radical change in psycho-analytic theory that libido is not pleasure-seeking but object-seeking. In connection with this, he holds, moreover, that erotogenic zones do not primarily determine libidinal aims, but are channels mediating object-relationships, and further, that a satisfactory theory of ego development must be framed in terms of relationships with objects and, in particular, those that have been internalised. I would re-formulate this in a more restricted way, as follows :

- (i) Since patients in the analytic situation are largely deprived of gratifications, they show an almost inexhaustible urge to develop new object-relations to the analyst.
- (ii) Compared with this urge, the rôle of erotogenic zones in the analytic situation is very small.
- (iii) The patient's development in the analytic situation is dominated by his internalisation of the analyst.

I have given my reasons elsewhere for modifying the formulation in this way, and can only summarise them here.

(a) Freud had difficulty in finding a suitable word for describing the intensity of an urge. Since 'Lust' in German does not have the appropriate sexual meaning, he introduced the Latin 'libido'. But the use of this word has become extended by analysts, so as to lose much of its original grossly sexual meaning ; it has become somewhat hazy. If 'lust' in English did not have the overtone of sinfulness, it would have served admirably. But it has not been used by Freud's translators. My point here is that if they had used it Fairbairn could not have developed this view, at least in its present form, because it would be impossible to say 'lust is not pleasure-seeking'.

(b) I do not deny the great importance of object-relations ; I only wish to point out that to exclude everything else is one-sided. The analytic situation is essentially one of object-relationship between patient and analyst. We must, therefore, allow for the analyst's 'parallax', i.e. that he employs a method that precludes him from observing non-object relationship phenomena. In trying to give a developmental theory of the human mind, based on transference phenomena, what is usually omitted is an admission that clinically observed behaviour is moulded by the framework of the analytic situation and by the counter-transference of the analyst. Thus the fallacy arises of regarding transference phenomena as a product of the patient alone (a one-person psychology) rather than as a product of the patient-analyst

combination (an object-relation or two-person psychology). In connection with this it should be mentioned that analysis is carried out in a state of abstinence, i.e. the patient is denied most gratifications during analytic sessions ; and it is surely significant that, while our knowledge of object-relations has greatly increased, our theories of gratifications have hardly developed since Freud and Ferenczi. Further it should be questioned whether the invariable qualities that we tend to ascribe to the human mind, on the basis of transference observations and interpretations, may not be the product of a standard analytic technique.

It cannot be said that I underrate the importance of the two-person relationship, for I have stressed, ever since 1932, that an original one-person situation (as assumed in the idea of primary narcissism) is a myth. I argue only that the tools of our trade do not permit us to infer that there are no non-object-relationship tendencies or behaviour.

I would therefore agree with Fairbairn if he would allow the qualification that his theory holds only so far as the analytic situation and his own individual technique go. Thus, while regarding object-relations as of fundamental importance, I do not agree that pleasure-seeking should be excluded.

The fundamental problem now arises of finding the relationship between these two tendencies of the libido : (a) for the patient in analysis and (b) in the development of the human mind. There would seem to be the following possibilities :

- (i) Pleasure-seeking and object-seeking are both innate and independent.
- (ii) Part of pleasure-seeking becomes at some stage transformed into object-seeking.
- (iii) Pleasure-seeking is a special case of object-seeking when the object-choice is a matter of indifference (i.e. when any one of a whole host of objects will do).

I do not think it is possible to decide in favour of one or other of these solely on the basis of clinical experience. If the question can be decided at all, it will have to be by other means.

MICHAEL BALINT

Comments on Fairbairn's Paper

It is good that independent thinkers in our midst should from time to time examine our key concepts, so that we remain alive to the progress of human thought in science and philosophy, and thus able to revise what in psychoanalysis is transient and tied to certain epochs, without however losing

COMMENTS ON FAIRBAIRN'S PAPER

sight of what remains true and is permanent. It is in this spirit that the following comments on Ronald Fairbairn's interesting paper are offered.

As to Fairbairn's critique, I agree with the author on some principal points, but for the sake of discussion will here concentrate on the others.

In contrast to other current and related schools of thought in psycho-analysis, Fairbairn at least makes clear that his formulations are incompatible with some of Freud's fundamental concepts, such as the analytic concepts of instinct, the libido theory, and mental topography, not to speak of 'minor' matters such as repression.

1 *Instinct*

It is widely recognised that the application of this concept to the human mind is full of problems. It appears that much which in former times was designated as *instinct* is in fact learned, complex behaviour, learned as a result of object relationships which in turn have become internalised. Social psychologists among others (cf. Newcomb¹) have pointed this out convincingly.

For Freud *instinct* is a borderline concept, having both a physico-chemical and a mental aspect. His view is basically a materialistic one which permeates the whole libido theory. The *id* also is conceived as a reservoir of instinctual energy, the source of which lies in the body, as well as a force presenting the mental apparatus with drive. It would appear that a fundamentally organismic, unified theory would be more correct, eliminating for instance the unnecessary dualism between the mind and the body.

On the other hand, Freud's concept is that of *Trieb*, impulsive drive, and does not accurately correspond to the term instinct. Freud himself called the doctrine of instincts 'our mythology'. Nowhere is his almost poetical notion in this respect more clearly in evidence than in the concept of Life and Death Instinct. The problems of the 'Death Instinct' are well known and have been exercising the minds of psycho-analysts since the inception of this concept. These problems are particularly interesting in the light of entropy and the second law of thermodynamics. The basic objection remains that in psycho-analysis physical concepts are being applied, either directly or via biology, to a medium for which they are inadequate, that of human interaction, i.e. psychology.

2 *Libido Theory*

First of all a remark on Fairbairn's objections to Freud's hedonism. It would appear that he takes this hedonism too seriously, and thinks too much of the person instead of a principle of regulation. It is true that Freud also spoke of 'Man the eternal seeker after lust', but here as always one can quote from his rich and concentrated writings in support of either

¹ Theodore M. Newcomb, *Social Psychology*, London, 1952

view. Freud's work must be taken as a whole to do it justice. However, there is no doubt that Freud, in a more scientific vein, thought of the *pleasure principle* more as of a regulatory tendency in the instinct household itself, as a tendency towards the reduction of tension, not as a pleasure-seeking device. The reality principle is not underrated with Freud. He always conceived a conflict between the pleasure and the reality principle.

A more important objection here is that *energy* is considered in Freud's theory apart from structure. The two ideas are conceptually distinguishable, and in psycho-analysis we find the separation of energy from structure useful. Such all-important concepts as for instance *fixation*, *displacement*, *sublimation*, or *withdrawal of libido* are dependent on that abstraction. One need only think of the metapsychological differences between hysteria and schizophrenia in terms of *loss of cathexis* in inner object representation to show how helpful and useful these abstractions are to us.

Actually, in Freud's later theory, as is well known, *libido* in the sense of *libido sexualis* became replaced by a view which takes into account the whole, the total economy of Life and Death. Here again it is true that *life* and *death* are symbols, mythological concepts, abstractions, and so are life—and death—instincts. But these have a bearing on clinical observation and are useful for our theoretical orientation. The present writer for one finds that his technical power has become much enhanced, and the therapeutic results correspondingly improved, since he has given due weight to self-destructive drives and their organisation in the super-ego (as distinct from the ego ideal).

Thus, while there are difficulties about the concepts of instinct and of libido, the present writer finds that they are technically indispensable. In this respect he disagrees with Fairbairn.

3 Object Relations

The emphasis on object relations is the most significant move within recent years; it comes from some quarters amongst psycho-analysts (and also amongst medical psychologists generally). Fairbairn has the merit of being one of the psycho-analysts who has given much thought to this, and done much work in this respect over a long period of time.

Szasz states in a recent paper:¹

In current psycho-analytic thinking, there is general agreement that psycho-analysis is first and foremost concerned with the study of *object relationships*. The problem of the 'economy' of object relationships is clearly not the same as the economic problem of instinctual energy.

¹ Thomas S. Szasz, 'Entropy, Organisation and the Problem of the Economy of Human Relationships', *International Journal of Psycho-Analysis*, 1955, 36, Parts 4 and 5, 289

COMMENTS ON FAIRBAIRN'S PAPER

Later Szasz writes :

The second problem is the need for new theoretical concepts regarding the economy of object relationships, based not on notions borrowed from physics, but on the operational method of psycho-analysis.

This is reflected in the increasing tendency to consider the transference/counter-transference relationship as of central importance in the psycho-analytic process.

The present writer feels that an even better field for the study of these interpersonal relationships is the analysis of small groups. He thinks that many of the problems with which Fairbairn is concerned can be solved, if we keep clearly in mind whether the relevant concepts refer to:

- (a) one person, conceived as an isolated unit (here belong the original psycho-analytic notions of 'instinct', libido, mental economy, etc.)
- (b) two-person relationship (the psycho-analytic situation, transference, counter-transference, resistance, etc.)
- (c) relationship between three people (the writer's 'model of three', the smallest model of a group)
- (d) multipersonal or group situation (the writer's 'group specific factors' and many concepts from group-analysis belong here).

The basic situation—also historically the oldest—is the group or community situation, which should be the matrix from which to define the position of the isolated individual or the two-personal transference relationship, and not, as so often happens, the other way round, when an attempt is made to 'explain' group-dynamics in terms of transference.

This view was also expressed by Freud, but he did not always hold consistently to it. In the last resort such a step implies that psycho-analysis would not any more belong exclusively to the natural sciences, but to the social sciences as well.

The present writer particularly agrees with Fairbairn and others about man's social nature, man being primarily a social animal. The partly explicit, but more often implicit, assumption inherent in psycho-analysis that social drives, social needs are secondary derivatives leads to many mistakes and fallacies.

Szasz, whom I will quote again and with whom I agree, makes it clear that these two sets of concepts, the biological and the social ones, are not—as Fairbairn thinks—mutually exclusive or incompatible.

I would like to emphasize in this connection that the considerations put forward in this paper have nothing to do with 'disproving' the classical 'economic point of view' of psycho-analysis. What is valid in one frame of reference does not become invalid or useless with the introduction of a new or different frame of reference. Instead it becomes necessary that we pay attention to the nature of the phenomena which we are interested in understanding better: appropriate theoretical frameworks must then be found for different problems.

Accordingly, Freud's economic concepts may or may not be helpful, depending upon the nature of our approach to a problem.¹

The pressing problem in this connection is to review the concept of primary narcissism.

4 Structure

It is not possible here to go in any detail into Freud's concepts of *ego*, *id*, and *super-ego*, but it may be said that these are constructions, abstracted from the living organism, which must be perceived as a whole in action. Possibly Fairbairn has something similar in mind when speaking of dynamic structures.

The introduction of the term 'id' made it possible to differentiate a mental province which, though unconscious, was not unconscious for dynamic reasons, not repressed. At the same time it became possible to say that part of the ego itself, in particular the super-ego, is unconscious. Thus the antithesis of the conscious and the 'unconscious' was replaced by the ego and the id. Wherever there is id the primary process reigns. The super-ego, as Fairbairn rightly suspects, is part of the ego, but is at the same time unconscious, and in so far is in close communication with the id. The 'ego' represents the structured aspect of the 'id'.

There is no implication of a conflict between mental structure and mental energy, and therefore Fairbairn's arguments, based on this assumption, are not valid. The *ego* is originally the mental representation of reality and the body, but even in the early days Freud ascribed to the ego particular instincts, the so-called *ego-instincts*, so that conflict was always between id-impulse and reality, or id-impulse and ego-impulse, not between impulse and structure. The social situation, the interpersonal relationship, the object relationship are particularly represented by their precipitates in the super-ego, thus allowing for Fairbairn's object-seeking characteristics.

However, it is not possible here to do justice either to the way in which classical psycho-analytical theory conceives the relationship between id, ego, and super-ego, nor to Fairbairn's positive contribution. His views have led to fruitful discussion among analysts and I am in agreement with some of them.

To end this paper the present writer may be permitted to say a few words about the help philosophy of science could give to a psycho-analyst.

I feel it would be a task for the philosophy of science to examine how far a science like dynamic psychology or psycho-analysis needs new criteria for the evaluation of its results and observations, distinct from those of physics, and how these could be embodied into an overall structure of science.

¹ Thomas S. Szasz, 'Entropy, Organisation and the Problem of the Economy of Human Relationships', *International Journal of Psycho-Analysis*. 1955, 36, Parts 4 and 5, 289

SOME COMMENTS ON DR FAIRBAIRN'S PAPER

It seems to the present writer that the concept of *science* might have to be changed, so as to do justice to a dynamic psychology which is based on the social nature of man, on the inter-personal nature of the data, on the fact that the human observer and the observed—interacting—provide the elementary data for our theory.

It would seem that there is a question to be decided whether science has absolute standards, valid for ever, or whether science is not itself a *tool* which has to change with the changing need of the human community, and if so, in what way this should be adjusted now to the new relative facts with which we psycho-analysts, amongst others, have to deal.

S. H. FOULKES

Some Comments on Dr Fairbairn's Paper

FREUD's metapsychological theories were, as he repeatedly stressed, his speculations on how the phenomena of psychological conflict could be explained keeping in mind not only his clinical findings, but also his knowledge of the central nervous system and of contemporary biological thought. The most striking clinical fact he had to contend with was the intensity of the pressure with which impulses or ideas, discovered to be sexual in nature as a rule, sought discharge against the wishes of the individual. The discharge of these foreign elements characteristically relieved tension, usually accompanied by a feeling of gratification. He therefore thought he had to account for drives or forces of a peculiarly impelling kind in these impulses and for the repression of their origins and nature from consciousness. The final forms his theories took were his concepts of the id, the ego, and the super-ego. Freud included the instincts in the id, but they were not regarded as giving the id an organisation. The id was 'a chaos, a seething cauldron of excitement' which acted as a reservoir of energy. Instincts for him did have objects and aims, nevertheless they were so plastic in man that what seemed to matter most for him was the 'drive' in them. He therefore retained as the sources of this drive the libido or sexual energy and the destructive energy of the death instinct, both of these psychic energies being derived from transformation of the activities of body cells.

In relating energy to structures in the psychic apparatus in a way that would account for the specific manifestations of psychological conflicts and behaviour, Freud had eventually to create the dynamic structures of the ego and the super-ego. At birth there was only the id with its plastic instincts determining relatively unstructured behaviour. Within the next few years there developed from it as a result of the experiences of the child the two dynamic structures, the ego and the super-ego, and the energy for their

functions was taken over respectively from 'desexualised' libido and the death instinct.

As Fairbairn indicates, these concepts present difficulties which can be looked at from two angles, namely, their consistency with the general scientific and biological thought of today and their value in clinical work. These two angles are naturally closely related yet, as in other branches of science, gaps and even inconsistencies between 'pure' and 'applied' theory can exist until certain facts demand changes in one or both. Fairbairn states that changes are now demanded on both grounds.

On the general grounds, it would be widely accepted that the notion of a free energy with an aim is out of step with the views of energy in the physical sciences; and from general biological considerations it seems unlikely that the constellation of instinct patterns in man would be so relatively unorganised as is assumed in the id concept and the libido theory. But, as Fairbairn points out, it is the unsatisfactory picture of behaviour given by the libido theory that is more important. It implies that conflicts arise from the need to seek pleasure, an affect which Freud related to the conditions under which id tensions were discharged. There is here, however, a gap between 'pure' metapsychological theory and working clinical hypotheses, and because the level of abstraction of the libido concept is so different from that at which the psycho-analyst works, the difficulty has been largely ignored. At the clinical level, the phenomena of conflict between conscious and unconscious motives are in fact interpreted largely in terms of object relationships. It is incompatibility between the urges to do certain actions with objects, and in particular the conflict between loving and hating relationships, that constitutes the core of analytical work. Thus, most psycho-analysts would accept Fairbairn's view that conflict is essentially related to interpersonal factors. In other words, in accounting for the nature of conflict at the clinical level Freud's divorce between energy and structure hardly enters into consideration. It is, for example, the libidinal or the aggressive impulses *towards the mother or father* that matter rather than the operation of the hypothetical libido or death instinct. Indeed many analysts do not accept the theory of a death instinct. Instead, their instinct theory resembles one which would be more in keeping with instinct theory in animals, i.e. it involves impulses to specific actions with objects. At the same time, it would not be true to say that the present instinct theory of the id is not without its influence on clinical work and here Fairbairn's plea for a more systematic object relations theory, i.e. one in which all drives are conceived as having specific aims with objects, is in my view justified not only in the interests of logical tidiness, but also to alter concepts which may be impeding advances in theory and practice. The fundamental nature of the issues he raises makes it difficult to do justice to his views in a brief statement, and the following points are merely a few of the considerations his views have suggested to me.

SOME COMMENTS ON DR FAIRBAIRN'S PAPER

(i) The large bulk of evidence regarding the nature of neurotic conflict shows increasingly that unconscious motives are unconsciously sought relationships ; they represent desired relationships with internal objects which the individual wishes to achieve in the external world. The work of Mrs Klein and her associates, Fairbairn, and others has shown that analysis as an instrument of investigating unconscious activity can be increased greatly in power and scope by the use of a more systematic object-relations theory. Thus, if we take the phenomena of auto-erotism and narcissism, these were formerly, under the influence of the libido theory, regarded as precluding analysis when present in an intense degree because they represented object-less discharge of libido. Mrs Klein's work, however, has shown that even severe schizophrenics with markedly narcissistic withdrawal can be kept in an analytic relationship by the constant interpretation of their behaviour in terms of relationships with objects of the most primitive kind. A meaning can be given to the auto-erotic and narcissistic behaviour when it is regarded as deriving from the earliest internalised objects, the nipple and breast, and such interpretations permit the subject, at least for a period, to make more externally directed relationships.

(ii) An object-relations theory is the most useful one in accounting for the most characteristic phenomena of the analytic relationship, namely transference, and for the effects of interpretations. There is a widespread agreement that it is chiefly when the individual experiences towards the analyst the aims of his unconsciously sought relationships that changes in the intensity of these occur and hence fundamental changes in the personality. Yet the full implications of this well known observation have not been adequately related to the psycho-analytic theory of the personality.

(iii) A libido type of theory, i.e. one in which drive is regarded as divorced from structure in its origins, is still used by some psychiatrists as the basis of their clinical action, but the clinical findings do not confirm the value of the theory. Thus it is on the basis of this type of theory that changing the balance of the sexual hormones, or other biochemical processes, is thought to hold out a means of altering some of the sexual perversions, because through such somatic interference it is hoped to change the libidinal tension. No substantial or consistent changes, however, in the pattern of sexual behaviour have been obtained in this way.

(iv) The great lack in the theory of personality at present is a framework with which the chief determinants of behaviour can be formulated and assessed. The identification of the main unconsciously sought relations would give a good account of the forms of relationship with people and things which the individual seeks in the external world, and with clearer identification, it should be possible to make even rough quantitative assessments of the strength of these relationships. With such a scheme, better predictions than we are able to make at present could be achieved, especially in

regard to the course of events to be expected during analytic treatment where they could be tested. A description of the personality in such terms would also permit of more precise meaning being given to such notions as 'ego-strength'. Hitherto analyses of very disturbed patients have at times had to be given up, or have not been undertaken, on the grounds that the ego was too weak for the unconscious forces the patient had to control. Concepts of ego strength, however, have never been very satisfactory in relation to clinical work and general theory. An object relations theory can make more specific what is missing, usually the structures derived from the normal good experiences of the child. When such relations with good internalised objects are too impoverished, there arises the problem of whether or not the analysis can proceed without some replacement experience. A few analysts, e.g. Winnicott and several American analysts, believe that such patients need to be permitted to make a relationship with the analyst in which the latter can be felt more directly as a person, e.g. by his being in limited ways more 'active'. Most analysts at present do not share this view, and only further clinical work with more adequate means of assessing changes in the dynamics of the personality will throw light on this issue.

Fairbairn's case for a psychology of the personality in terms of its object relations is therefore in my view important and timely. His formulations have clearly sharpened his clinical observations and these have been widely regarded by many psycho-analysts as constituting an important contribution. With regard to his particular psychic structures, it is perhaps premature to comment on their value as such views require considerable testing out in practice. It is interesting to note, however, that Winnicott's independent formulations of a 'true self', i.e. a self that can give free expression to feeling, and a 'false self', i.e. a self that conforms to the inhibiting pressure of the outer environment, closely resemble Fairbairn's structures.

While the libidinal and antilibidinal egos may represent the outcome of the attempt of the personality to reach maximum integration, these structures give the impression of a greater degree of organisation than appears to be the case, at least in some patients. If we take the antilibidinal ego in many severe hysterics, clinical findings would suggest that this has active sub-structures. Thus in order to secure libidinal gratification, such persons may identify predominantly with one parent and be persecuted by the attacks of the imago of the other, and then at other times reverse the predominance of the identification. Interestingly enough too, a common remark of such patients is that they do really know 'who they are', i.e. they have no feeling of a stable constellation in either of these larger structures of Fairbairn's.

A last point may be made concerning metapsychology and internal objects. There can be little doubt that it was the inadequate nature of the means at his disposal for relating energy to structure that prevented Freud

REPLY TO BALINT, FOULKES, AND SUTHERLAND

from solving this problem more to his satisfaction. Had he had available the concepts of modern physics and neuro-physiology, he would have been freed from the restricting effects of what Colby¹ describes as his 'hydraulic metaphors'. When the cathexis energy of ideas, etc. in the psychic apparatus can be related, as Colby has done, to the frequency period, synchrony and dysynchrony of pulsations, then the structural nature of drive can be much more readily illustrated by a model. As a prelude to the description of his model, Colby makes almost the same criticism of the classical id, ego, and super-ego theory as does Fairbairn. A critical point is that he considers that drive structures must incorporate the notion of purposive aims and the image of an internal object must be closely related to the aim of a drive schema. The development of Colby's ideas will therefore be of considerable theoretical interest. He is careful to point out that his model merely serves at present as a basis for relating energy and structure, and that there is inevitably a considerable gap between his metapsychological theories and clinical application. Nevertheless, and this is Fairbairn's main thesis, the more the theories at these different levels are in harmony, the better it must be for clinical progress.

J. D. SUTHERLAND

Fairbairn's Reply to the Comments of Balint, Foulkes, and Sutherland

THE Comments of Balint, Foulkes, and Sutherland upon the views expressed in my critical evaluation of certain psycho-analytical conceptions would seem to indicate, if nothing else, that there are other psycho-analysts besides myself who are not altogether satisfied with the classic formulations of psycho-analytical theory. It is obviously impossible for me, within the limits of a brief reply, to deal in any detail with the many issues raised in these Comments; and I must therefore restrict my reply to the submission of a few considerations of a somewhat general nature.

It must be admitted as a general principle that the conditions under which data are obtained have an influence upon the form assumed by the data themselves, and that consequently a certain relativity attaches to all data and to such theories as may be based upon them. It is to this principle that Balint appeals when he submits that my views only apply in relation to (a) the analytical situation in general, and (b) my individual technique in particular. This submission applies equally, of course, to all psycho-analytical theories whatsoever; and, in my opinion, Balint presses his submission to a point at which psycho-analytical theory becomes reduced to a state of such subjectivism as to leave no alternative but an attitude of Humean scepticism. Such an attitude would, of course, be incompatible

¹ K. M. Colby, *Energy and Structure in Psychoanalysis*, New York, 1955

with the aim of science, which is to establish explanatory principles possessing a universal objective validity. That the aim of science is itself a limited aim based upon limited values (viz. purely explanatory values), and that the practice of psycho-analysis as a therapeutic measure is necessarily influenced by other human values which preclude the possibility of the analytical session conforming to the rigorous requirements of an experimental situation in the generally accepted sense I should be the first to maintain;¹ and, since one of the implications of this point of view is that science is merely an intellectual tool, and therefore not so much a determinant as a servant of other values, it would appear to be in conformity with an estimate of the rôle of science somewhat tentatively suggested by Foulkes. It must be recognised, however, that, in the capacity of a tool, science can be of value only on the assumption that its findings possess at least approximately universal validity. It is to such validity that psycho-analytical theory must aspire; and both Ezriel² and I myself³ have tried to show that the nature of the analytical situation by no means precludes the fulfilment of such an aspiration. So far as concerns the principle to which Balint appeals, however, it is my opinion that the conditions of the orthodox analytical situation (involving as they do isolation of the patient in a recumbent position on the couch, and an attitude of detachment on the part of an analyst who is invisible to the patient) have an influence very different from that which Balint supposes. It must be recognised that, since the patient, *qua* patient, may be presumed to have suffered from severe deprivations in childhood, he comes to the analytical situation with an intense craving for object-relations already present in him, and that, since the conditions of the orthodox analytical situation impose upon him a severe deprivation of object-relations with the analyst, they have the effect of reproducing the trauma of deprivation from which he originally suffered. An artifact is thus introduced into the observed data. But, contrary to Balint's contention, the effect of the artificially induced trauma is to compromise such capacity for object-relations as the patient possesses, to provoke in him actively the 'regressive' phenomena to which Winnicott has drawn attention (as Sutherland notes), and to compel him to fall back upon the pleasure principle and the primary process as defensive techniques.⁴

¹ W. R. D. Fairbairn, 'Observations in Defence of the Object-Relations Theory of the Personality', *Brit. J. Med. Psychol.*, 1955, 28, 154-156

² H. Ezriel, 'The Scientific Testing of Psycho-Analytic Findings and Theory', *Brit. J. Med. Psychol.*, 1951, 24, 30-34

³ W. R. D. Fairbairn, 'Theoretical and Experimental Aspects of Psycho-Analysis', *Brit. J. Med. Psychol.*, 1952, 25, 122-127

⁴ I have recently been led to the conclusion that, so far from constituting basic forms of psychic activity, the pleasure principle and the primary process essentially represent defensive techniques of a non-specific character, as implied in the text above.

REPLY TO BALINT, FOULKES, AND SUTHERLAND

The effect of the orthodox psycho-analytical method is thus to confer an exaggerated importance, not upon object-seeking phenomena, but upon phenomena of a pleasure-seeking nature. This limitation does not, of course, apply to the situation involved in the analysis of small groups, to which Foulkes makes reference.

Balint's etymological diversion regarding the respective meanings of the German 'Lust' and the English 'lust' can hardly be allowed to pass without brief reference in view of his statement that, if Freud's English translators had used the term 'lust' instead of 'libido', I could not possibly have propounded the view that libido is object-seeking; but all that it seems necessary to remark by way of reply is that my views are concerned with concepts rather than with the terms used to describe them. It is, however, relevant to point out that the formulation to which Balint refers is one which appears in a paper originally published in 1941,¹ and that, in the light of the subsequent development of my views, I should now prefer to say that it is the *individual in his libidinal capacity* (and not libido) that is object-seeking. This reformulation is designed to avoid any appearance of that hypostatisation of instincts which is criticised in the foregoing paper. It should perhaps be added that there is no question of my denying the importance of the rôle played by pleasure in the mental economy. What is at issue is the particular rôle which it plays; and my contention would be that, whilst there can be no doubt that under certain conditions it can become an 'end', its natural function is that of a 'means'.

Importance must be attached to the fact noted by Foulkes that 'for Freud *instinct* is a borderline concept, having both a physico-chemical and a mental aspect'. Freud was, of course, a neurophysiologist before he became a psychologist; and, although the modern science of psychopathology represents the harvest of Freud's insight into the fact that the phenomena of mental disease can only be satisfactorily understood and explained in psychological terms, he himself never abandoned the lingering hope that these phenomena would eventually prove capable of explanation on a biochemical basis. This duality of outlook (not to mention ambivalence) must be regarded as having had the effect of introducing unsatisfactory features into a number of Freud's concepts—and conspicuously those of the pleasure principle, the instincts, and the id as an element in the mental constitution. Thus his lingering neurophysiological bias led him to treat the sources of psychical energy as lying outside the psyche, and to conceive the 'id' in terms which render its psychical status dubious. Herein lies the ultimate significance of my criticism of that separation of energy from structure which it is the aim of my concept of 'dynamic structure' to overcome. If psychology is to be taken seriously as an

¹ W. R. D. Fairbairn, *Psychoanalytic Studies of the Personality*, London, 1952, pp. 31-32

explanatory system, it must be assumed that psychical energy is inherent in the psyche. It must also be assumed that the psyche is a structure in which this energy is inherent—and not, as is implied in Freud's views, that psychical structure (in the form of 'the ego') is the product of the mutual friction of biochemical energies arising within the organism and environmental agencies. Foulkes quotes Szasz in support of his view that there is no real incompatibility between biological and social concepts and that these may be fruitfully combined; but it would be my contention that it is fatal to clarity of thought to introduce into one science explanatory principles belonging to another. It is the inherent aim of psychology to explain human behaviour and experience in strictly mental terms; and, if this aim is to be fulfilled, the concepts employed in psychological explanations must be exclusively psychological. In so far as Freud is consistent, his description of unconscious mental processes conforms to this requirement; but many of his concepts are not properly psychological at all. As Foulkes points out, some of them, e.g. his concept of the instincts, actually embody a mythological component; but it is more common for his psychological conceptions to be adulterated with biological or biochemical components—his conception of the 'actual neuroses' being a case in point. Further, it seems to me a postulate of psychology as an independent science that the proper subject of psychological investigation is not the organism, but the *person*; and it was under the influence of this opinion that on a previous occasion I expressed the view that 'as in the case of all forms of psychological research, the investigations of psycho-analysis should be conducted at the level of personality and personal relations'.¹ My theory of dynamic structure and my elaboration of an explicit object-relations theory of the personality represent an attempt to implement this view. The latter theory was, of course, really initiated by Freud himself in his conception of the super-ego; and Melanie Klein's conception of internal objects represents a further step in the development of this theory. On the other hand, there is little evidence of the influence of the conception of dynamic structure in Freud's formulations; for, after all, he abandoned his theory of 'ego instincts', which Foulkes cites as an example of such an influence, in favour of the view that the ego is assailed by impulses from outside the psyche. Thus arose the anomaly of an ego whose impulses are inherently alien to it, and of instincts essentially alien to the personality to which they belong. And indeed it may even be said that the 'ego' described by Freud is really a façade-ego dependent for its existence upon repression and other defences. Sutherland has appositely referred to Winnicott's recent colloquial formulations of 'the false self' and 'the true self'.² It was found by Winnicott

¹ Fairbairn, 'Observations . . .', as cited, 151

² D. W. Winnicott, 'Metapsychological and Clinical Aspects of Regression within the Psycho-Analytical Set-Up', *International Journal of Psycho-Analysis*, 1955, 36, 16-26

REPLY TO BALINT, FOULKES, AND SUTHERLAND

that, in the cases which he has described, 'the false self' had to be dissolved away (with a resulting regression) before 'the true self' could with difficulty emerge; and similar cases are familiar in my experience. 'The false self' corresponds, of course, to Freud's 'ego'; but 'the true self' is a structure for which no place can be found in Freud's theory of the mental constitution which Winnicott has accepted without modification. At the same time, as Sutherland has pointed out, 'the true self' of Winnicott's formulation corresponds closely to what is described as 'the libidinal ego' in the alternative theory of the mental constitution which I have submitted¹—and in accordance with which, it may be added, an 'impulse' is psychologically meaningless except as a manifestation of activity on the part of an ego-structure. For the rest, all that can be said on the present occasion is that, whilst Foulkes finds the abstract concept of 'energy' useful on the grounds that such explanatory concepts as 'fixation', 'displacement', 'sublimation', and 'withdrawal of libido' depend upon it, my view is that such concepts can be reformulated with greatly enhanced significance in terms of the theory of dynamic structure and the object-relations theory of the personality. It is also my view that the significance of the self-destructive phenomena, to which Foulkes rightly attaches such importance, can be best understood, not in terms of the abstract 'drives' to which he refers, but in terms of active attacks directed by one internal ego-structure against another in the manner which I have already described elsewhere.²

A final note in reply to Sutherland's comment on the ego-structures envisaged in the theory of the mental constitution which I have submitted as a substitute for Freud's. According to Sutherland, the libidinal and antilibidinal egos which I describe 'give the impression of a greater degree of organisation than appears to be the case, at least in some patients'; and he contends that, in the case of advanced hysterics for example, the antilibidinal ego would appear to be composed of (or at any rate contain) several active sub-structures. In support of this contention, he cites the reversal of rôles which may occur in cases in which identification with one parent in a libidinal rôle is accompanied by internal persecution on the part of the imago of the other parent. It should be noted, however, that in this instance there is no question of the *antilibidinal ego* reversing its rôle, which is always that of a persecutor. What is changed is the internal object with which the antilibidinal ego is identified. The example quoted by Sutherland would thus be better calculated to support a contention that it is the internal objects postulated in my theory that are loosely organised and must contain sub-structures. Actually I find no difficulty in accepting the proposition that the internal objects are composite structures; and

¹ W. R. D. Fairbairn, *Psychoanalytic Studies of the Personality*, London, 1952, pp. 94-119; and 'Observations on the Nature of Hysterical States', *Brit. J. Med. Psychol.*, 1954, 27, 107-109

² *Ibid.*

indeed it would be my contention that this is so. Thus the internal objects which I envisage may be composed of maternal and paternal components in all proportions and in all degrees of integration; and, for that matter, they may undergo both disintegrative changes under pathogenic conditions and integrative changes under therapeutic conditions. I find it difficult to believe, however, that, except in cases of advanced schizophrenia, the disintegration of internal objects often reaches a point at which my differentiation between the exciting, the rejecting, and the ideal objects (if only as constellations) becomes meaningless. As contrasted with these internal objects, the three ego structures which I describe would seem, characteristically, to be much more definitely organised and differentiated—a phenomenon which may be attributed to the simple fact that they are ego-structures, in contrast to object-structures. Indeed, in the light of clinical experience I find it difficult to imagine any structure more obdurately encapsulated in its own organisation than the antilibidinal ego; and it is to this fact more than any other that we must look for an explanation of the intensity of the resistance, which is so characteristically present even in the most favourable subjects for analysis. Doubtless the libidinal ego is, as a general rule, less definitely organised than the antilibidinal ego. Nevertheless the obstinacy with which it clings to its exciting object and to its chosen modes of gratification indicates a tightness and rigidity of organisation which is only excelled by that of the antilibidinal ego, and which makes a formidable contribution to the resistance. It is in the central ego that it is commonest to find the looseness of organisation to which Sutherland refers; and it is to this source that one must attribute the remarks of the advanced hysterics whom he quotes as complaining that they do not know who they are. Actually in these cases there is often comparatively little of the central ego left except the function which it exercises in repressing the libidinal and antilibidinal egos; and, since this repression is, after all, only a negative function, it is small wonder if such patients are assailed with doubts regarding their identity.

W. RONALD D. FAIRBAIRN

Experimentation within the Psycho-Analytic Session

FEW readers of Dr Ezriel's article are likely to dispute his view that the hypotheses of psycho-analysis have not been exposed to systematic test, and should be. Dr Ezriel's own suggestions are welcome, even if, as I think, he minimises the difficulties involved in clinical experimentation.

I shall confine my discussion to two questions: (i) Why is the practising analyst unlike an archaeologist or historian? (ii) Can systematic experiments, designed to test specific hypotheses, be carried out by analysts?

(i) One reason proposed by Dr Ezriel for rejecting the archaeological analogy is that many of the ostensible recollections of patients turn out to be fantasies, or distortions of some past experience. So what the analyst finds out about is not the past history of the patient, but his present unconscious state.

This does not seem sufficient reason for abandoning the historical viewpoint in psycho-analysis. If it were sufficient, there would be no historians (or no rational historians). Historical evidence includes forgeries, fantasies, deliberate and undeliberate distortions, as well as some moderately accurate records. No-one believes that uncritical faith in the reliability of his sources is the mark of the competent historian. If the psycho-analyst wishes or thinks it therapeutically necessary to play the part of a historian, he will not let the difficulties inherent in all historical enquiries stop him. Owing to the privileged position of the psycho-analyst *vis-à-vis* his subject of enquiry, who is also his main source, his historical task is in some ways easier, though in some ways more difficult, than that of the normal historian or biographer. Even if most of the ostensible recollections of their early lives by patients are fantasies or distortions, it requires a historical enquiry to show that this is so.

Another of Dr Ezriel's anti-historical arguments is that whereas the archaeologist has to draw his inferences about the past from such relics as happen to have survived and to have come to his notice, the analyst deals with material unconsciously selected by his patient as significant and belonging together, material which is offered spontaneously or in response to the analyst's intervention.

Leaving aside the caricature of the trained archaeologist as a haphazard spademan, and the fact that most biographers and historians have to deal with material selected (consciously or unconsciously) by their subjects as significant and hanging together, the important point here seems to be that the analyst is in a specially privileged position, since (on Freud's theory of transference) the past history of the patient is being partially re-enacted in his present attitudes and behaviour towards the analyst. But this surely should help the historically-minded analyst, not hinder him. A partial analogy would be if a biographer (not an analyst) secured an interview with his subject, wishing to find out more about that person's attitude to politicians, opposition to whom had played an important part in his subject's life. If the biographer were himself a practising politician, or brought one along to the interview, he might hope that his subject would in his behaviour during the interview provide some useful clues, whatever the value of his deliberate disclosures about the past.

The most important difference between the clinical analyst and the historian is that the former's primary aim is practical—the treatment and, if possible, cure of his patients. If he tries to unearth his patients' infancies, it is because he believes this will be of therapeutic value. The historian seldom

has any analogous practical aims to which his historical studies are subordinate ; if he has, we think of him as a reformer, or propagandist, or as an ' applied historian ', lacking the ' disinterestedness ' of the pure historian. It is not even true of most historians that they try to explain how present-day conditions have been brought about, as Dr Ezriel suggests. This exaggerates the likeness between psycho-analyst and historian. The analyst who became fascinated in his patient's past and ignored his present state would be guilty of professional neglect ; no such reproach could be made against the historian, in his professional capacity.

The clinical analyst with theoretical interests in the testing of hypotheses has to subordinate his theoretical to his practical aims, in case of conflict. These theoretical interests can be distinguished from those of the historian, who makes his inferences about past events *with the aid of* general principles, but not with a view to testing them. He uses these principles as a means to the discovery of particular historic truths. Conversely, the theoretical analyst may try to confirm his hypotheses by showing that they make possible successful retrodiction : he is interested in such particular retrodictions simply as means to the testing of his theories, not as ends in themselves, unlike the historian. Thus, both historian and scientist may rejoice at the confirmation of some particular retrodiction, but the reasons for their joy will differ. The one is glad that a historical speculation has been confirmed, the other, that a general hypothesis has been strengthened. The same person may rejoice both scientifically and historically, but his joys will be logically distinct.

How much credit can the hypotheses of psycho-analysis derive from successful inferences about the patient's past ? Not much, in my view. If a hypothesis can be strengthened by such successes, it can be weakened by failures, and it would be difficult to collect statistical evidence of the relative frequency of successes, and to compare this with the success-frequency of retrodictions made by other methods. In any case, failures could always be explained away plausibly, and the principles used would not be risked in any way. It would be necessary to ascertain precisely which hypotheses were used ; hypotheses not used are not directly tested, and a successful retrodiction could not be treated as a blanket confirmation of all psycho-analytic hypotheses. The possibility that alien, non-analytic principles had been used would require investigation. It would be difficult to formulate rules on what was to count as a success ; e.g. the degree of precision required, and the strength of evidence required for the verification of the historical statement. Striking retrodictive successes might be allowed to have ' suggestive ' value, but would not, I think, give support to hypotheses for which there was not strong evidence obtained in other ways. I believe this agrees substantially with Dr Ezriel's views, to whose discussion of clinical experimentation I now turn.

PSYCHO-ANALYTIC SESSION EXPERIMENTATION

In considering objections to his experimental programme, Dr Ezriel does not discuss the restrictions imposed by the medical status of the analyst. To the layman, absence of experimental animals appears a considerable handicap since the analyst is unable to vary the initial conditions in ways that are thought likely to be therapeutically harmful, or retarding. Dr Ezriel maintains that clinical experimentation depends on the analyst's ability to control the antecedent variables, but this suggestion of experimental freedom appears misleading. The control of the analyst consists apparently in his correct identification of the patient's 'unconscious object-relations' and the incorporation of these in his interpretation. If these are the limits of his control he has not much greater freedom in varying the initial conditions than the person asked to attend an identification parade and to pick out his assailant. A wrong identification may of course be made, but if the analyst aims—as he surely must—at a correct interpretation, he is not a free man. Therapeutic obligations are theoretical fetters.

Dr Ezriel's suggestions for the experimental testing of the two laws he formulates deserve closer study than I can here give, and my discussion will be tentative owing to my fear of having misunderstood them. The laws are logically complex, being compounded out of empirical and theoretical concepts, with bridges from one to the other called 'operational rules'. Their full clarification would raise many methodological issues crucial for the whole system of Freudian theory and practice.

As regards the testing of these laws, which are of the form if *P* then *Q*, test-possibilities arise only if the antecedent conditions are realised. For instance, in the second law the antecedent conditions include by implication the stipulation that the analyst should have correctly inferred 'the hidden dynamics of the patient-analyst relationship'. Only if this condition is satisfied can the second law be tested.¹ The criterion for the fulfilment of this condition is that certain operational rules have been correctly applied by the analyst to the material presented to him. For the purposes of testing this law, the adequacy of these rules is assumed, for without this assumption there is no means of ascertaining whether the initial conditions necessary for the testing of the law are fulfilled (or whether the observed consequences do in fact conform to the predicted consequences). Since these rules are being thus used in the testing of the law, they are not themselves being tested, and confirmation of the law by successful prediction does nothing to confirm *them* (or of course the correctness of the analyst's interpretation, which is decided by reference to these rules). But these rules connecting behaviour-statements with 'transcendent' statements seem basic to Freudian theory,

¹ This appears inconsistent with Dr Ezriel's statement that the correctness of the analyst's interpretations, while essential for successful therapy, is irrelevant for theoretical purposes. Perhaps he refers here to other hypotheses, the testing of which does not presuppose a correct interpretation.

and experiments which assume their validity might be thought to have rather low theoretical importance. Dr Ezriel's second law could be regarded as a practical prescription or rule of therapy, of the form 'If the analyst wishes to bring about such-and-such a change in his patient, let him act in such-and-such a way'. In no field is it more difficult to distinguish between practical rules and fundamental theory than in psycho-analysis, but this does not mean that no attempt should be made to distinguish them, and to clarify the relationship between theory and practice.

Another worrying point concerns the falsifiability of hypotheses. What is to count as a falsification? Dr Ezriel states that if the response predicted does not occur, it is legitimate to assume the existence of unknown additional factors, provided that in some cases such predictions are successful. This implies that a hypothesis which has been confirmed on 'some' occasions of its use is thereby rendered immune from the possibility of future falsification.

It is obviously difficult to state all the conditions relevant to the testing of a psycho-analytic hypothesis in the course of a session. Dr Ezriel does not include the analyst's personality and flair among the relevant conditions, and refers to him as a detached and passive observer, except at times when he is giving his interpretations. Is not his earlier characterisation of the situation as one in which two human beings are interacting with each other, more realistic?

To conclude this discussion (which ignores much of Dr Ezriel's material), it might be suggested that a more promising approach to the testing of psycho-analytic theories lies in the attempt to generalise them and relate them to the conceptual schemes of general psychology. This would be good for both parties, by fertilising the one, and offering a wider testing-range to the other. Freud himself would surely have approved.

R. J. SPILSBURY

University College
Aberystwyth

*Reply to Mr Spilsbury*¹

FOR reasons of space I shall deal only with the main points raised by Mr Spilsbury.

(1) My suggestion that the exclusive use of 'here-and-now' *interpretations*² allows experimentation within the psycho-analytic session does not,

¹ *Erratum*: In my paper in the Freud Centenary Number, 'Experimentation within the Psycho-Analytic Session', 1956, 7, p. 36, line 3, *for* historical *read* ahistorical.

² The term 'here and now' was introduced into psycho-analytic literature by J. Rickman to emphasise that the psycho-analytic method is essentially an ahistorical one concerned with forces operating between patient and analyst within the analytic

REPLY TO MR SPILSBURY

as Mr Spilsbury appears to infer, entail abandoning the historical *viewpoint* in psycho-analysis. The historical viewpoint is clearly essential with regard to aetiology, i.e. in genetic studies of factors which shape the child's developing personality. But in fact 'here-and-now' interpretations do justice to the historical viewpoint, in that—as opposed to psychotherapeutic comments which emphasise the 'there and now', the patient's *current* life outside his sessions—they deal with the unconscious precipitates in the patient's mind of his unresolved infantile conflicts, as they manifest themselves in interactions with the analyst in which the patient aims at or avoids certain relationships. The need to retain the historical viewpoint is therefore not at issue; the real question is what kind of *interpretation* given to the patient is most useful therapeutically and for the understanding of the dynamics of the particular personality under investigation or of human behaviour in general, i.e. whether 'here-and-now' interpretations are adequate means for these purposes or whether it is necessary also to include explicit references to the patient's past—in particular, attempts at historical reconstructions of events of his childhood.

As regards analytic *therapy*, there is general agreement among analysts that the most important factor is the interpretation of the relationship between patient and analyst. If, nevertheless, historical references have so far been retained as well, this may be simply because they have always been used since they were first employed by Freud before he discovered transference, and their utility and dynamic significance have never been seriously examined. However, my recorded sessions, e.g. the one quoted in my paper, indicate that the exclusive use of here-and-now interpretations is at least as effective as the customary mixture of historical and transference interpretations. It is noteworthy that in 1937 Freud wrote about historical reconstructions: ¹

Often enough we do not succeed in bringing the patient to recollect what has been repressed. Instead of that, if the analysis is carried out correctly, we produce in him an assured conviction of the truth of the construction which achieves the

session. Subsequently, in putting forward hypotheses about the nature of these forces—as manifested in the material produced by the patient—I have used the phrase "“ here-and-now ” interpretations" to describe interpretations which point out the here and now existing patient-analyst relationship in terms of the three object relationships connected by a 'because' clause, as described in papers I published in *Psychiatry* (1952) and in this *Journal* (May, 1956). Dr J. O. Wisdom in a review in the same number of the latter journal (p. 108) also uses the term "“ here-and-now ” interpretations", not, however, in the sense attached to it by me, but (Wisdom: personal communication) 'to denote any interpretation making some kind of reference to here and now forces that excludes the extra-analytic situation'.

¹ S. Freud, 'Constructions in Analysis', *Int. J. Psycho-Anal.*, 1938, **19**, 385. In this paper Freud refers to historical reconstructions as the analyst's 'work of construction, or, if it is preferred, of reconstruction . . .' (379).

same therapeutic result as a recaptured memory. The problem of what the circumstances are in which this occurs and of how it is possible that what appears to be an incomplete substitute should nevertheless produce a complete result—all of this is material for a later enquiry.

I think this statement not only indicates the hypothetical nature of such reconstructions of the past but is also in keeping with the view that historical comments by the analyst, and what seem to be extra-transference responses of patients to them, may really be disguised here-and-now interactions. Such comments might therefore be not only unnecessary but at times even disturbing to the analytic relationship in so far as they are felt by the patient not as interpretations but as *actions* of the analyst. Sometimes they may increase the patient's anxiety through the analyst's apparently deliberate evasion of the patient's most urgent problem, or his failure to understand it ; or may postpone its resolution by providing a reassuring escape route from here-and-now tensions.

My main reason at present for using here-and-now interpretations is, however, the *scientific* one that, as opposed to historical reconstructions, they allow experimental work *within the setting of the psycho-analytic session*, in the way described in my paper. Whether the analyst wishes to be an historian or not, psychical reality and transference make it impossible for him to get reliable historical evidence from the material produced by the patient within the session. I think Mr Spilsbury overlooks that whereas an historian may have ways of sifting out the truth when faced with contradictory evidence from different sources, the analyst has to deal with contradictory statements from one source only, i.e. the patient, and can uncover only the currently operating unconscious object relationships which motivate them. Thus while he can build up a picture of the patient's Unconscious as it exists at the time of the analysis—and in the case of a child patient can also study the developmental changes taking place during the period of treatment—his investigation does not enable him to reconstruct an objective picture of the patient's past ; it can yield hunches, and perhaps reveal in what succession certain layers of this patient's personality were formed, but not the dates at which they were established or the events that caused them.¹

If, nevertheless, the analyst wishes, for any reason, to act as an historian, he may of course look to *extra-sessional* sources—e.g. a diary kept by the

¹ Although many analysts hesitate to admit these limitations of the psycho-analytic method regarding historical reconstructions of childhood events from the analysis of *adult* patients, the majority seem to recognise them in regard to child analysis when they reject (I think rightly) Melanie Klein's claim of being able to produce from the analysis of older children something more than merely hypothetical reconstructions of the mental life of infants during the first year of life. (Cf. E. Glover, 'An Examination of the Klein System of Child Psychology' in *Psychoanal. Stud. Child*, 1945, I, pp. 75-118, or E. Kris, 'Psychoanalytic Propositions' in *Psychological Theory*, ed. M. H. Marx, New York, 1954, p. 341.)

REPLY TO MR SPILSBURY

patient's parents at the time—for evidence to confirm or disprove his historical reconstruction. But quite apart from the frequent impossibility of getting any such reliable evidence, there is the further difficulty that, even where it appears to have been obtained, it does not necessarily follow that there is a *causal link* with the material uncovered in the sessions. For example, it is not sufficient for an adult patient to produce a 'memory' of having been seduced or having witnessed parental intercourse as a child, and for objective extra-sessional evidence to be obtained that this did in fact happen, since exactly the same kind of 'memory' may be uncovered in cases where it can be definitely shown that this event could not have happened to the patient in material reality. The 'memory' could be accepted as being a recollection of the real event only if it included such *specific details*, confirmed by the extra-sessional evidence, as to rule out the possibility of its being a fantasy or pseudo-memory which (according to the patient's present unconscious needs) may weave in memory-bits taken from different contexts and times. It is not often that such stringent conditions are fulfilled.

(2) Mr Spilsbury says that most biographers and historians have to deal with material selected, consciously or unconsciously, by their subjects as significant and hanging together. I am not sure in what sense this statement is to be understood (e.g. how it could apply to the survival of archaeological relics), but any such 'selection' is quite different from the patient's unconscious but purposive selection of material *for the analyst* to whom he comes for treatment. It is this difference which explains why, whereas it is in general a matter of chance whether an historian gets sufficient data to explain a particular episode, it is not a matter of chance that the patient in need does produce for the clinician all the required dynamically significant elements, as a result of his unconscious desire to establish with the clinician a relationship of a kind determined by those elements.

(3) On Mr Spilsbury's discussion of clinical experimentation, I should like to reply to three points.

One is his criticism that I refer to the analyst as a detached and passive observer although, according to my own description, the analytic situation consists of two human beings interacting with one another. The point here is that the better analysed the analyst, the more detached and passive will he remain (giving interpretations being ideally his only active step) and the more clearly will there emerge the patient's contribution to this interaction, i.e. that aspect of his personality which is dominant in that session and around which the experiment can revolve.

Another point that has worried Mr Spilsbury is my supposed attempt to make my 'second law' immune from the possibility of future falsification. Here I would emphasise, first, that, as I indicated in my paper, a study of my recorded case material to date (including both here-and-now interpretations and other kinds given as controls) yields no example of the phenomenon

described in the second law taking place after other than here-and-now interpretations based on the first law and on the operational rules which I use. Secondly, this phenomenon can be correctly predicted sufficiently frequently, and sufficiently specifically, to rule out the possibility of its being due to chance. Thirdly, although in principle alternative explanations may be possible, none has yet been produced which similarly enables successful predictions to be made ; unless and until such an alternative explanation is produced, which accounts for at least as many of the observed facts, e.g. not only for the cases where my second law is borne out but also for some of those where it does not appear to be, the various propositions outlined in my paper (including the assumption of as yet unknown factors in those instances where the second law does not appear to hold) must, I think, be accepted as, at the minimum, together constituting the best available working hypothesis—which is the respectable status of most scientific theories in general use.

A third point raised by Mr Spilsbury concerns the operational rules. He suggests that even if their application leads to successful predictions which confirm the second law, this would do nothing to confirm the rules themselves. If I understand this point correctly, it leads to the question whether the analyst is merely applying some set of arbitrary rules to the pre-interpretation and post-interpretation material in such a way that he sees whatever he expects or wishes to see ; in particular, as regards the application of these rules to the post-interpretation material, the problem is whether the analyst has any objective means of testing whether his prediction has been fulfilled. I should therefore emphasise in the first place that the rules are not subjective speculations which will vary with each analyst, but that, though as yet not as precise as one could wish, they show a fair degree of objectivity and, having been defined beforehand, they can be applied by different trained observers who, especially when dealing with less complex (i.e. with unambiguous) material, ought to arrive at the same conclusion independently. Further, in a certain number of cases it happens, as in the session quoted in my paper, that the ideational content of the avoided relationship becomes *completely* manifest after the interpretation, so that there is no need to apply any of the operational rules to the post-interpretation material ; here the correctness of the prediction can be seen directly. In addition, sessions such as that quoted in the paper contain material which allows the direct observation, first of unconsciously operating mechanisms which distort the thoughts the patient wishes to express, and then of the application to these thoughts of corresponding operational rules that undo the distortion. For example, the impulse to use a vulgar word for passing wind, through the mechanisms of 'displacement' and 'turning into the opposite', is transformed into the 'nice smell of hyacinths' ; the application in the interpretation of the appropriate operational rules to the 'nice smell' discloses behind it the originally

RELATION OF PHYSICAL LAWS TO ORGANISMS

avoided word. Finally, the operational rules are not merely, as it were, an instrument used to *study* the pre-interpretation and post-interpretation material (i.e. to observe the situation before and after the experimenter's manipulation), but actually determine the analyst's *manipulative intervention* in the experiment, since he applies them to the pre-interpretation material in order to identify the three object relationships defined in the first law, as the essential first step in formulating his interpretation. I therefore think that the application of the operational rules is part of, and not merely something preceding, the testing of the second law, and I cannot agree with Mr Spilisbury that the successful predictions which are made possible only by the application of these rules do not imply a confirmation of the rules themselves.

(4) I do not think Mr Spilisbury is correct in suggesting that the analyst's experimental freedom is hampered because he has an obligation always to aim at a correct interpretation. There is considerable dispute among analysts as to the kind of comment that should be made to the patient, and for this reason alone the effects of a wide variety of types of interpretation could be, and ought to be, studied systematically. Quite apart from this, I only wish that the frequency of correct interpretations were yet great enough for the obligation to aim at them to be felt as a burdensome restriction on experimental freedom. As opposed to what happens in many other kinds of experimental situation, the psycho-analytic experimenter cannot work out his intervention in advance but has to decide it by analysing, on the instant, the material produced by the patient. Subsequent close study of the sessional material, by means of recordings, will provide sufficient examples of the absence of a correct interpretation to satisfy the most rigorous statistician seeking controls. This was one of my reasons for saying in my paper that, while the correctness of the analyst's interpretations is essential for successful therapy, the incorrect interpretation is equally interesting for the study of the patient's behaviour. My main reason was, however, the possibility of testing hypotheses concerning the effects of interventions by the analyst which are not correct interpretations.

HENRY EZRIEL

On the Relation of Physical Laws to the Processes of Organisms

It has been suggested that we already know the basic laws applicable to structures in organisms : non-relativistic quantum mechanics.

This is ambiguous. It can mean (A) quantum mechanics can predict ¹

¹ By defining the initial conditions under which organic systems and processes (a) develop, and (b) persist, and assuming the empirical values of any dimensionless constants required.

the characteristic processes of organisms, or (B) the processes of organisms are compatible with quantum mechanics. (B) is weak, may be true, but cannot yet be confirmed. (A) is at present untrue, and (I shall argue) will probably remain so. Moreover quantum mechanics is not a completed theory.

There are reasons for believing that if the complex stereo-structure of the processes, for example, in the functional parts of living cells, here called bio-complexes (such as regions in which replication occurs, enzyme systems, polarisable and contractile membranes, etc.) is to be suitably represented, quantum mechanics must be supplemented by additional structural principles. This is a reasonable expectation since it is unlikely that quantum mechanics is the final word in the theory of basic structures of different orders of complexity. For this it is, in a technical sense, too weak.

Quantum mechanics uses an approximative procedure¹ of a special kind based on the correspondence principle, by which eigenvalues are expressed as power series ($\sum A_n \cdot \epsilon^n$) in certain development parameters ϵ (whose values are empirically but not yet theoretically determined), Schrödinger's equation corresponding to the first variable term of such a series. The dependence of quantum mechanics on this 'perturbation' procedure is the result of its having treated as separable variables (e.g. those associated with protons, electrons, and photons) which are mixed in relativistic terms and coupling terms representing interactions of these particles. The efficacy of the method depends² on the smallness of the basic ϵ , i.e. the fine-structure constant (α) and the ratio of electron to proton mass (β), and also of the effective number (N) of the degrees of freedom. Quantum mechanics can make no unambiguous predictions when N is too large, nor could it even for N small if α and β were not small compared with unity.

This quantum mechanics procedure has provided excellent approximations for a vast range of systems where the energy is low and N small. For example, crystals can be treated because, though complex, they are ordered in a known manner, so that N is low. For other complex systems (nuclei, particle showers, superconductors, bio-complexes) the form of ordering is not yet known, N is too large, no satisfactory theory has been found, and new methods, stronger than those of quantum mechanics, may be necessary. Quantum mechanics can treat more kinds of complex systems (e.g. atoms under the Pauli exclusion principle) than could classical mechanics, but not all kinds. Thus we cannot yet say to what extent the methods of quantum mechanics can be used to predict the conditions of formation and the characteristic properties of bio-complexes, and in so doing to throw light on

¹ N. Bohr, *Journal of Chemical Society*, 1932, I, 349

² Bohr, *loc. cit.* (for α). M. Born and J. R. Oppenheimer, *Ann. d. Phys.*, 1927, 84, 457 (for β), where the ϵ is the fourth root of the ratio of electron to mean nuclear mass.

RELATION OF PHYSICAL LAWS TO ORGANISMS

the spatial and temporal co-ordination of parts and their processes in a single developing and adaptive organism.

One view is that this should ultimately be possible because (a) the energies are small and the 'relativistic effects' therefore (?) negligible and quantum mechanics sufficiently accurate; and (b) increasing complexity introduces no new principles (?) and the technical difficulties of computation will gradually be overcome. This argument is misleading. 'Relativistic effects' (in the widest sense of higher terms involving c , whose systematic treatment lies outside the scope of quantum mechanics) may be non-negligible even at low energies, e.g. in spin and other quantum interactions, and the deeper difficulties are of selection, not computation. In the expansion for the energy the coefficients of the successive terms representing these effects increase rapidly when N is not small. Thus not only high energy, but also high N , may prevent the early terms from decreasing sufficiently rapidly to provide good approximations, so that the method loses all value. It may be possible to try other series, but there is no method of finding one which provides good approximations. The approximative procedure of quantum mechanics appears therefore to be essentially inappropriate to complex systems where N is large.

Thus Schrödinger's (non-relativistic) equation by itself leads to definite predictions only when N is small, so that the form of the particular solution to be used is obvious, e.g. in the theory of molecules the shape of the nuclear skeleton on which the ψ function is to be draped. The number of possibly appropriate solutions increases rapidly with N and the equation soon loses all predictive power, unless supplemented by additional information, drawn not from quantum mechanics but from experience, which keeps N low enough. Even the determination of the ground state of the H_2 molecule requires the exclusion of several alternatives by trial, not by any general rule. The step from this to the transformations of, say, an enzyme system (probably involving non-classical orientation properties) is so immense that new ordering principles reducing N (like the Pauli principle and spin coupling rule, but more powerful) are necessary to restrict the myriad possible solutions to the class which actually occurs. The failure lies not with the mathematical techniques but with the physical principles which no longer provide an appropriate method of selection and approximation.

The quantum chemist has now to select the stablest arrangements of minimal energy by trial and error guided by experience and tentative rules of symmetry, and physical theory would be strengthened if rules could be found directly relating the degree of stability to the character of a structural pattern. In the organic realm the precise equilibrium structures of proteins, nucleic acids, etc., are now being experimentally determined, and in a generation their exact deformations and transformations in course of

function may also be known. If theory is to keep pace it will require principles showing how certain classes of arrangements of atoms are established and maintained in organisms, i.e. how these are discriminated from other possible arrangements of the same atoms. The present argument suggests that *these principles of organic ordering cannot possibly be deduced from quantum mechanics alone*, but that they must be established independently by observation and then absorbed into a comprehensive theory of complex structures and their transformations.

Quantum mechanics is in certain respects a consistent and unified theory. But it is neither empirically final, nor theoretically complete. It leaves open many problems in particle theory as possible growing points: nuclei, mesons, relativistic formulations, coupling theory, many-body potentials, low temperature properties, relation to gravitation, etc. My conclusion is that new developments (such as a more powerful relativistic theory of many body potentials or the equivalent) must lead to the discovery of more general structural principles and more powerful approximative methods *before* physical theory will be able to provide a satisfactory representation of the principles characterising organic processes. Here the main issue is: What symbolic method is best suited to represent the functional cycles and transformations of structure of bio-molecular complexes and in so doing to reveal in a satisfying manner the reason for the unique rôle of proteins and nucleic acids in organisms?

L. L. WHYTE

A Second Note on Degree of Confirmation

I The suggestion has been made by Professor J. G. Kemeny¹ (with a reference to my definition of *content*), and independently by Mr C. L. Hamblin² that the *content* of x , denoted by ' $C(x)$ ', should be measured by $-\log_2 P(x)$ instead of $1 - P(x)$, as I originally suggested. (I am here using my own symbols.³) If this suggestion is adopted, then my *desiderata*³, for *degree of confirmation* of x by y , denoted by $C(x, y)$, have to be slightly

¹ John G. Kemeny, *Journal of Symbolic Logic*, 1953, 18, p. 297. (Kemeny's reference is to my *Logik der Forschung*.)

² C. L. Hamblin, 'Language and the Theory of Information', a thesis submitted to the University of London in May 1955 (unpublished); see p. 62. Mr Hamblin produced this definition independently of Professor Kemeny's paper (to which he refers in his thesis).

³ 'Degree of Confirmation', this *Journal*, 1954, 5, 143 *sqq.*; see also pp. 334 and 359.

A SECOND NOTE ON DEGREE OF CONFIRMATION

amended : in (ii) and in (v), we must replace ± 1 by $\pm \infty$; and (iii) becomes :

$$(iii) \quad 0 \leq C(x, xy) = C(x, x) = C(x) = -\log_2 P(x) \leq +\infty.$$

The other desiderata remain as they were.

Mr Hamblin suggests ¹ to define degree of confirmation by

$$C(x, y) = \log_2(P(xy)/P(x)P(y)) \quad (1)$$

which for finite systems, but not necessarily for infinite systems, is the same as

$$C(x, y) = \log_2(P(y, x)/P(y)) \quad (2)$$

a formula which has the advantage of remaining determinate even if $P(x) = 0$, as may be the case if x is a universal theory. The corresponding relativised formula would be

$$C(x, y, z) = \log_2(P(y, xz)/P(y, z)). \quad (3)$$

The definition (1) does not, however, satisfy the desideratum viii (c), as Mr Hamblin observes ; and the same holds for (2) and (3). *Desiderata* ix (b) and (c) are also not satisfied.

Now the desideratum viii (c) marks in my opinion the difference between a measure of explanatory power and one of confirmation. The former may be symmetrical in x and y , the latter not. For let y follow from x (and support x) and let a be unconfirmed by y . In this case it does not seem satisfactory to say that ax is always as well confirmed by y as is x . (But there does not seem to be any reason why ax and x should not have the same explanatory power with respect to y , since y is completely explained by both.) This is why I feel that viii(c) should not be dropped.

Thus I prefer to look upon (2) and (3) as highly adequate definitions of *explanatory power*—of $E(x, y)$ and $E(x, y, z)$ —rather than of degree of confirmation. The latter may be defined, on the basis of explanatory power, in many different ways so as to satisfy viii(c). One way is as follows (I think that better ways may be found) :

$$C(x, y) = E(x, y)/(1 + nP(x)P(\bar{x}, y)) \quad (4)$$

$$C(x, y, z) = E(x, y, z)/(1 + nP(x, z)P(\bar{x}, yz)) \quad (5)$$

Here we may choose $n > 1$. And if we wish viii(c) to have a marked effect, we can make n a large number.

In case x is a universal theory with $P(x) = 0$ and y is empirical evidence, the difference between E and C disappears, as in my original definitions, and as demanded by desideratum (vi). It also disappears if x follows from y .

¹ 'Degree of Confirmation', this *Journal*, 1954, 5, p. 83. A similar suggestion (without, however, specifying 2 as basis of the logarithm) is made in I. J. Good's review of my 'Degree of Confirmation'; cf. *Mathematical Review*, 1955, 16, 376

Thus at least part of the advantages of operating with a logarithmic measure remain : as explained by Mr Hamblin, the concept defined by (1) becomes closely related to the fundamental idea of information theory. Mr Good also comments on this point (see footnote 4).

The transition from the old to the new definitions is order preserving. (This holds also for explanatory power, as Mr Hamblin's observations imply.) Thus the difference is metrical only.

2 The definitions of explanatory power, and even more of degree of confirmation (or corroboration or acceptability or attestation, or whatever name may be chosen for it) give of course full weight to the *weight of evidence* (or the 'weight of an argument' as Keynes called it in his chapter vi). This becomes obvious with the new definitions, based upon Mr Hamblin's suggestions, which seem to have considerable advantages if we are at all interested in metrical questions.

3 However, we must realise that the metric of our C will depend entirely upon the metric of P . *But there cannot be a satisfactory metric of P , that is to say, of logical probability.* To show this we consider the logical probability of any measurable physical property (non-discreet random variable) such as length. to take the simplest example. We make the assumption (favourable to our opponents) that we are given some logically necessary finite lower and upper limits, l and u , to its values. Assume that we are given a distribution function for the logical probability of this property ; for example, a generalised equidistribution function between l and u . We may discover that an empirically desirable change of our theories leads to a non-linear correction of the *measure* of our physical property (based, say, on the Paris meter). Then our 'logical probability' has to be corrected also ; which shows that its metric depends upon our empirical knowledge, and that it cannot be defined *a priori*. In other words, the metric of the 'logical probability' of a measurable property would depend upon the metric of the measurable property itself ; and since this latter is liable to correction on the basis of empirical theories, there can be no purely 'logical' measure of probability.

These difficulties can be largely, but only partly, alleviated by making use of our 'background knowledge' z . But they establish, I think, the significance of the topological approach to the problem of both degree of confirmation and logical probability.

But, even if we were to discard all metric considerations, we should still adhere, I believe, to the concept of probability, as defined, implicitly, by the usual axiom systems for probability. These retain their full significance, exactly as metrical geometry retains its significance even though we may not be able to define a yardstick. This is especially important in view of the need to *identify logical independence with probabilistic independence* (special multiplication theorem). If we assume a

A SECOND NOTE ON DEGREE OF CONFIRMATION

language such as Kemeny's (which, however, breaks down for continuous properties) or a language with *relative-atomic* statements (as indicated in Appendix 1 of my *Logic of Scientific Discovery*), then we shall have to postulate independence for the atomic, or relative-atomic, sentences (of course, as far as they are not 'logically dependent', in Kemeny's sense). On the basis of a probabilistic theory of induction, it then turns out that we cannot learn if we identify logical and probabilistic independence in the way here described; but *we can learn* very well in the sense of my *C*-functions; that is to say, we can corroborate our theories.

Two further points may be mentioned in this connection.

4 The first point is this. On the basis of my axiom systems for relative probability ¹, $P(x, y)$ can be considered as defined for any value of x and y , including such values for which $P(y) = 0$. More especially, in the logical interpretation of the system, whenever x follows from y , $P(x, y) = 1$, even if $P(y) = 0$. There is thus no reason to doubt that our definition works for languages containing both singular statements and universal laws, even if all the latter have all zero probability, as is the case, for example, if we employ Kemeny's measure function m , by postulating $P(x) = m(x)$. (In the case of our definitions of E and C , there is no need whatever for departing from the assignment of equal weight to the 'models' (cf. Kemeny, op. cit. p. 307). On the contrary, any such departure must be considered as a deviation from a *logical* interpretation, since it would violate the equality of logical and probabilistic independence demanded in 3, above.)

5 The second point is this. Among the derived desiderata, the following are not satisfied by all definitions of ' x is confirmed by y ' which have been proposed by other authors. It might therefore be mentioned separately as a tenth desideratum: ²

x. If x is confirmed or corroborated or supported by y , so that $C(x, y) > 0$, then (a) \bar{x} is always undermined by y , i.e. $C(\bar{x}, y) < 0$, and (b) x is always undermined by \bar{y} , i.e. $C(x, \bar{y}) < 0$.

It seems to me clear that this desideratum is an indispensable adequacy condition, and that any proposed definition which does not satisfy it is intuitively paradoxical.

KARL R. POPPER

University of London

¹ This *Journal*, 6, p. 56 sq. (see also pp. 176 and 351). Simplified versions are given in *British Philosophy in the Mid-century* (ed. by C. A. Mace), p. 191; and in my *Logic of Scientific Discovery*, Appendix * iv

² Compare the remark in this *Journal*, 1954, 5, end of the first paragraph on p. 144.

REVIEWS

Foundations of Quantum-Mechanics: A Study in Continuity and Symmetry.
By A. Landé.

Yale University Press, 1955. Pp. 106. 32s.

IN various articles ¹ as well as in a book Professor Landé has suggested a new approach to quantum mechanics. Whereas most presentations of the subject rest content with the assertion that one has to accept the peculiar features of the atomic world (as expressed, e.g. by the principle of complementarity, or by the commutation-rules, or by the calculus of the ψ -functions) as fundamental, however difficult it may be to understand them properly, Landé makes the attempt to elucidate those features by showing how they can be understood on the basis of some very simple and plausible assumptions. The assumptions are (a) Leibniz's principle of continuity (an infinitely small change of a cause cannot lead to a finite change of the effect) and (b) some further principles, to be mentioned presently.

Applying (a) to the case of two different diffusing gases (or else of two different states of one and the same gas) leads to the postulate that the maximal increase of entropy resulting from the mixture should be dependent on the degree of likeness of the components. This postulate, which Landé calls the 'postulate of entropy-continuity', removes a very curious paradox of statistical mechanics which was first discussed by Gibbs, and leads at the same time to some well known features of quantum mechanics: ordering of states into sets of mutually orthogonal states; splitting effect; jumps; the Born interpretation.

For example, let us start with the class K of all possible states of a certain system (gas) and choose a state A_1 , then a state A_2 completely unlike A_1 (i.e. a state which on mixture with A_1 gives the classical value for the maximal increase of entropy), then a state A_3 which is completely unlike A_1 as well as A_2 and so on until no further state is left which is completely unlike all the preceding states. In this way a class (A_i) of 'mutually orthogonal' states is obtained. Take from the remaining states a state B_1 and proceed as above until there is no further state left which is orthogonal to all the B_i already selected, which gives the class (B_i) and continue in this way until all members of K have been sorted out in one way or another. This method

¹ I only mention: 'The Logic of Quanta', this *Journal*, 1956, 6, 300; 'Quantum-Indeterminacy, A Consequence of Cause-Effect-Continuity', *Dialectica* 1954, 8, 199; 'Continuity, a Key to Quantum Mechanics', *Philosophy of Science*, 1953, 20, 101 ff. (cf. also the literature given in this last article).

(which is not necessarily unambiguous) terminates in a classification of the elements of K into sets of mutually orthogonal states.

Again, within thermodynamics the relations between states are usually defined by ideal experiments. Two states are unequal if and only if there is a semi-permeable membrane which allows us to separate them. Introducing likeness-fractions means admitting that there are intermediate cases between complete separation and complete absence of separation. Thus, if an A -filter is a device which passes the state A but completely rejects any state which is orthogonal to A , it is to be expected that a state $B \neq A$ will neither pass completely, nor be completely rejected. B is split into two parts, one passing, one rejected (one mol of B -gas is split into $q(A, B)$ mols passing and $1-q$ mols rejected. This leads to an operational definition of the q 's as passing fractions.).

Further, according to the definition of an A -filter given above, the part of the B -gas which has passed the filter will be in the state A : the B -gas reacts to the filter by jumping partly into the state A , partly into the state \bar{A} . And as in the case of a B -gas exposed to an A -filter the only splitting is into A and \bar{A} , it follows that $q(B, A) + q(B, \bar{A}) = 1$, or, elaborating, $\sum_i q(B, A_i) = 1$ the summation being extended over all elements of (A_i) . The symmetry of the q follows from very simple thermodynamic considerations. Adding to this account the assumption that any kind of substance consists of indivisible units we arrive at the following concerning probability: $q(A, B)$ is the probability that a particle belonging to a gas in the state B will pass an A -filter when exposed to it: 'The correct statistical interpretation (Born) follows immediately from the thermodynamic origin of quantum-theory' (p. 56).

Apart from the principle of entropy-continuity, some of whose consequences we have just been discussing Landé uses (b_1) the principle that there is a general law which allows us to derive $\{q(A, B)\}$ (i.e. the matrix with the elements $q(A_i, B_k)$) from $\{q(A, C)\}$ and $\{q(C, B)\}$, as well as (b_2) the postulate of a constant probability-density in phase-space (a postulate which Landé erroneously identifies with Liouville's theorem). The simplest, if not the only possible way, to satisfy (b_1) consists in assuming that $\psi(A, B) = \psi(A, C) \cdot \psi(C, B)$ where $\psi = \sqrt{q} \exp(i\phi)$ —i.e. in assuming a 'law of superposition' for probability-amplitudes. (b_2) leads directly to Born's commutation-rules. And this finishes Landé's account of general quantum-mechanics.

Trying to evaluate this account we must at once admit its methodological importance. We have here a well-ordered presentation which makes it easier to understand the principles on which the whole theory rests. But the specific principles employed suggest also a new way of looking at this theory which is much nearer to classical thought than the favourite house-philosophy of many physicists would have it nowadays. For it is commonly

assumed that these matters have once and forever been settled by what is known as the 'Copenhagen-Interpretation'.

As regards the Copenhagen-Interpretation, one ought to realise that it is not an interpretation in the usual sense in which one speaks, e.g. of the Born 'interpretation'. The Born interpretation is necessary and sufficient for applying what would otherwise be a purely mathematical formalism to reality, or (if the word 'reality' is disliked because of its apparently 'metaphysical' character) for tackling physical problems in terms of quantum theory. Any additional element, however interesting and fruitful it may be, is not thus necessary and it is therefore not forced upon us by physics, or by 'experience', or by any similar source, it is a bit of speculation which we have to judge according to its fruitfulness, its inner consistence as well as according to its elucidatory power. This speculative, or, if you like, metaphysical character of Bohr's interpretation is often overlooked by people who are in the habit of already seeing quantum theory in its light and who therefore think that either physics, or simple analysis of physical theories forces them to adopt Bohr's views. Admitting this implies that we are to a certain extent free to invent and to consider other 'metaphysical' interpretations. Of all the interpretations which have been suggested so far, Landé's attempt seems to be the one 'next to reality', i.e. it seems to be the attempt which goes least beyond the Born interpretation and which therefore should be most suitable to throw light upon quantum mechanics as it is. And whatever one may think about minor points, one must certainly admit that the curious and unsatisfactory features of the Bohr picture are completely absent from Landé's presentation. Quantum theory is much nearer to classical physics than one is usually inclined to assume. Having shown this is another important asset of Landé's undertaking. (As it has been doubted whether Landé's approach is tenable on thermodynamic grounds it may be useful to point out that the partial separability of states, advocated by Landé, corresponds to the well known fact that, within quantum mechanics, states are completely separable by semi-permeable walls if and only if they are orthogonal. It was von Neumann (*Mathematical Foundations of Quantum Mechanics*, Princeton, 1955, p. 370) who first connected this theorem with Gibbs' paradox, but without making this fact the starting point of a new presentation of the theory.)

Two critical points ought to be mentioned, one general and one specific. First of all the connection between the postulate of entropy continuity and the general principle of continuity itself is hardly as close as Landé would have it (viz. direct entailment). For there are numerous cases in which a straightforward application of the more general principle leads to undesirable consequences. For example, a mass-point is moving on a straight line l with constant speed. A point divides l into two parts, l_1 and l_2 . The mass-point's passing from l_1 to l_2 involves a discontinuity.

REVIEWS

Now in this case it would hardly be advisable to introduce degrees of 'being in l_1 ' and 'being in l_2 ' and thus to abandon the concept of a continuous movement altogether. For the undesirable discontinuity can easily be removed by using a different presentation of the process and by treating the empirical aspect within a theory of errors. This shows that, although the principle of continuity may be adopted as a guiding principle, it is not sufficient for deriving the entropy-continuity from classical statistical mechanics. Another remark is about the rôle which the wave-function plays in Landé's presentation. Landé thinks that there exists a form of quantum theory, admittedly very impracticable, which does not contain any probability-amplitudes but only probabilities and which is equivalent to the usual presentation by means of probability-amplitudes; i.e. he thinks that the probability amplitudes have no independent physical meaning. Considering the importance of the symmetry-properties of the wave-function as well as of Schrödinger's equation I doubt whether this account of the matter is correct. But the value of Landé's undertaking for those who want to become clear about physics—and this should include also the physicists—remains unimpaired by those minor difficulties.

P. K. FEYERABEND

Niels Bohr and the Development of Physics.

Edited by W. Pauli, L. Rosenfeld and V. Weisskopf.

Pergamon Press Ltd., London, 1955. Pp. vii + 195. 30s.

THIS volume contains ten essays on subjects of theoretical physics in honour of Niels Bohr's seventieth birthday, giving testimony to the influence of the Copenhagen School in moulding the outlook of a whole generation of physicists. It is the outlook of instrumentalism, using K. R. Popper's term, which asks us to accept empirical facts and principles such as duality and complementarity because they 'work' without being interested as to *why* they work. The question is, of course, whether the quantum principles are considered as 'last principles', or whether one hopes to reduce them to something still more fundamental. Thus, whereas the Pauli exclusion-rule was originally introduced as a 'principle', Pauli himself has always searched for reasons of a general and immediately acceptable kind (such as symmetry of past and future, and of positive and negative charge) which might be made responsible for the exclusiveness of particles with half-integral spin and for congregation of particles with integral spin. The present volume contains a further contribution by Pauli to this topic. Our only objection is against Pauli's term 'theoretical *a priori* arguments' because even symmetry postulates are not *a priori*, however plausible they may be. A similar task of reducing accepted empirical rules to plausible ground

REVIEWS

postulates will have to be carried out sooner or later for the whole of quantum theory.

Of particular interest is an essay on the interpretation of quantum theory by Heisenberg who defends the statistical view, called the Copenhagen interpretation in this Bohr Volume, against those who are dissatisfied either with the present formalism, i.e. with the physical content, or with the language, i.e. with the underlying philosophy of the quantum theory. Their opposition stems in Heisenberg's view mainly from the apprehension that the statistical interpretation may lead us too far away from the idea that there is an objective world in which things happen without the presence of observers. Heisenberg tries to alleviate this fear on semi-philosophical grounds by introducing a distinction between the 'objective' and the 'actual'. He points out that such a distinction, as contrasted to the combined 'objective reality' of classical physics, can already be made in the domain of the Second Law of thermodynamics. Here one can assign an *objective* temperature to a body upon the evidence of the average velocity of its particles, some of which may escape and be recorded. The exact instance of a recording cannot be predicted; before it occurs there is only a potentiality, more exactly a probability. The recording turns this potentiality into an *actuality*. Yet, if on grounds of a microphysical knowledge of the particles one could predict the actual recordings as to their individual time instants, then one could not assign an *objective* temperature to the body any more, since the concepts of temperature and entropy presuppose statistical disorder. Objective temperature and actual recordings are mutually exclusive though 'complementary'. Furthermore, every individual recording leads to a reduction of the probability function, and this reduction is not contained in the mathematical equations of motion of the particles; it constitutes a 'subjective' element, according to Heisenberg. A similar situation prevails in quantum theory: A particle emerging in a certain state from a certain instrument is characterised by a ψ -function representing various potentialities for the future behaviour of the particle in reaction to other instruments. The potentiality of this individually unpredictable behaviour, described by a ψ -function, is termed *objective* in contrast to the *actual* result of an individual test. The latter suddenly reduces the original ψ to a new ψ , and this reduction is not described by the mathematical wave equation; it is subjective. Again, an objective ψ -situation and an actual test are complementary.

It seems here that Heisenberg makes a rather strained use of the terms 'objective', 'actual', and 'subjective'. A potentiality represented by a ψ -function which, quoting Heisenberg, 'so to speak contains no physics at all', 'being completely abstract and incomprehensible', yet being 'completely objective, i.e. no longer contains features connected with the observers knowledge' might as well be termed subjective, since it only

REVIEWS

describes expectations in the *mind* of a physicist. On the other hand, the result of an 'actual' test is also 'objective' in Heisenberg's sense of being independent of the knowledge or presence of an observer. In contrast to the clear and simple mutual exclusiveness of an actual momentum measurement and an actual position measurement, the proposed complementarity between an objective situation (be it temperature or ψ -function) and an actual record of an individual event does not seem too attractive. We doubt whether it will convince the opponents of the statistical interpretation on philosophical grounds, if they have not grown convinced by Heisenberg's earlier pioneer work in physics. If one wishes to convert determinists not merely by microphysical evidence but on more fundamental grounds, one may point out the inconsistency of believing in determinism and in the Second Law of thermodynamics at the same time, since the latter clearly transcends the bounds of classical determinism. The quantum theory only proposes special rules for the metric of the probabilities, rules which ought to be deducible (and in this reviewer's private opinion are deducible) by simple 'theoretical *a priori* arguments' using Pauli's expression.

When singling out for criticism the one paper of Heisenberg, we cannot do justice to other essays in the Bohr Volume written by outstanding leaders in modern theoretical physics, particularly in nuclear and field theory.

ALFRED LANDÉ

Operational Philosophy: Integrating Knowledge and Action. By Anatol Rapoport.

Harper & Brothers, New York, 1953. Pp. xi + 258. \$3.75.

THIS is a courageous attempt to make (in the words of the Constitution and Rules of the Philosophy of Science Group) 'an approach through the various special sciences to the philosophy of science', and also to the philosophy of ethics and politics. Dr Rapoport gives us, in a clear and pleasant style, an account of the operational theory of knowledge, with its familiar emphasis on definition, its distinctions between meaning, truth and validity, and its translations of the ontological into the pragmatic, the moral into the empirical: for instance 'Is X real?' becomes 'How stable is the invariant to which X refers?', 'Which factor is the more fundamental?' becomes 'Which factor can be more easily manipulated?', 'What is Good?' becomes 'To what do people refer when they speak of the "Good"?' and so on.

This is, however, far from being the mixture exactly as before. There is no artificial simplification of the method of science in the interests of a phenomenalist philosophy. Dr Rapoport abandons the paralysing demand

REVIEWS

that every scientific statement shall have direct operational meaning, and takes full account of the function of models in science. He characterises scientific knowledge interestingly as the discovery of *invariants* within the empirical data, and attempts, more controversially, to extend this idea to the field of ethics, by claiming that the search for invariants within existing ethical systems will reveal man's basic and unchanging needs and will thus provide supracultural criteria for judging between the various ethical systems.

Dr Rapoport disarms criticism by his good-humoured and undogmatic manner. 'The only immutable value is the possibility of questioning all values' (presumably there is an 'ought' here which is not translatable into an 'is'), and Dr Rapoport does not exempt even operational philosophy from the questioning. The approach throughout is pragmatic: it is emphasised that the use of the operational method in ethics and political theory may be only a first step, and that this step is taken because the method of science now seems to be the best available. There is refreshing recognition of the difficulties inherent in applying it even in the social sciences. There is an interesting chapter on 'The Extensional Bargain', in which the possibility of communication in non-operational discourse is discussed: 'Before one has learned to speak, one must babble', and amid the babbling Dr Rapoport discerns the task of an autonomous philosophy.

An honest and stimulating book, admirably adapted to the gap between popular and technical works which the author wishes to fill. It leads one to hope that he will follow up his forays across the frontiers between operational philosophy and poetry, and consider whether there is not in the world of reciprocal relationship, which is that of social life as well as of poetry, something which after all escapes the subject-object categories of operational philosophy as he describes it. It has been necessary to relax the requirements of that philosophy since the days of logical positivism in the interests of the rationality of physical science; may not a more profound study of the methodology of the social sciences lead to still greater relaxations?

A final plea: When will writers of the analytic school take their examples of theological statements from reputable theological works, and consider them in their context? Can it be that they have not studied the writings of theologians with the care which they devote to those of scientists?

MARY B. HESSE

Ethical Judgment: The use of Science in Ethics. By Abraham Edel.
The Free Press, Illinois, 1955. Pp. 348. \$5.

THIS is a book on ethics and the use of science therein by an American academic. American academics, unlike their businessmen who tend to be rotund tranquil men, are haggard, anxious, active men, living in a ruthlessly

REVIEWS

competitive world in which production in bulk is both an obligation and a passion. They are financed by Foundations whose administrators want results, even if they don't understand them. Now all this is admirable. No realms yet unconquered by the human mind can surely resist such energy and optimism, so massive an onslaught so well financed and supplied by the rear echelons. Manifest destiny clearly wills that what is not yet known will fall to the professors.

But what happens in fields unsuited to this approach? What if in philosophy truths were few, simple, sad, and useless? (Can anyone imagine the prof, after his project, reporting to the Foundation in sentences that are few, simple, sad, and useless—or anyway openly so? Why, the very rules forbid it.) Philosophy may be more like climbing a mountain top, leading to a fine view but also to cold and fright and with nothing to bring back. Take the present subject, 'the use of science in ethics'. This presumably means the use of knowledge of facts in making choices and evaluations. Now is it likely that, substantially, anything more can be said on this matter than was said by David Hume, or that any merit attaches to saying it at greater length with less elegance? No, but if the professor were to believe and say this, there would be no lengthy book, no Foundation grant, perhaps no post. It is a pity Kierkegaard is not alive to utilise for the purposes of comic literature the theme of the 'project' that will tell us about Values, as he did so use the cottage industry (by comparison) of men like Hegel.

As with other works which make this big claim—to have discovered how to advance substantive ethics—one approaches with scepticism, but also with a guarded hope, for perhaps . . . after all . . . something has been discovered and the unplausible serpent really will teach us, at least a little, how to tell good from evil and be as gods are. Indeed the serpent claims he will. But not just yet; the next book, the next chapter, the next paragraph, for as yet there is this, that, or the other little academic difficulty to be overcome, allusion to be made, ambiguity to be clarified. (Great advances in substantive ethics are the kind of thing that is always due in the next chapter.) So patiently we read on, until—good heavens—we notice we must have *passed* the great revelation without noticing it, as a man who has spent the night awake in a train so as not to miss his station and then falls asleep just as he is entering it. So back we go to look for it (the index doesn't say where it is) and burrow deeper to find it.

When we have dug it up it turns out to be the discovery that as we come to know more about our wants and needs and the world in which we satisfy them, the amount of moral disagreement will diminish. Probably so, though it is also possible that the advance of knowledge and control will make our aims more indeterminate, or show up greater-than-expected incompatibilities of alternatives or interests.

REVIEWS

In addition to attaining this, the book contains lively discussions of how the diminution of indeterminacy (warranted moral disagreement, so to speak) does, or might, work with the help of various relevant sciences, and, at the beginning, a useful formulation of the problem of ethical truth as it may appear to those concerned with it for practical ends. Indeed, the central theme is not merely elaborated in philosophic abstraction: it is also pursued in detail in the context of the various human sciences. Biology, psychology, sociology (in a broad sense) and history each receive a chapter devoted to the discussion of their implication for ethics conceived as indicated above. These discussions are well-informed, and range wide, and their presence makes the book of interest to those specifically interested in the philosophy of science rather than philosophy in general. Each of these chapters would provide a good starting point for workers in the relevant field who wish to think systematically about the moral implications of their subjects, for the various themes which arise receive clear and forceful statement. Of its kind, the book is far, far better than most.

ERNEST GELLNER

Uncertainty in Economics and Other Reflections. By G. L. S. Shackle.
Cambridge University Press, 1955. Pp. xv + 267. 25s.

THIS book is a collection of eighteen articles which Professor Shackle has published between 1939 and 1953. The first hundred odd pages introduce, clarify, defend, re-state and extend the author's theory of expectations.¹ The opening essay in particular introduces the subject with an urbane forensic brilliance.

The next hundred odd pages contain contributions to the theories of interest, the multiplier, and of the macroscopic effects of taxation and government spending. There is a persistent emphasis on the future-consciousness, the delayed results, the hopes and fears, characteristic of economic activity.

The remaining fifty pages are on the philosophy of economics. The chief item here is a sketch of a classification of economic theories according to the rôle they accord to time and to expectations. First comes static economics, involving full information (and so no uncertainty about the future), rationality, and frictionless adjustments towards equilibrium. Next Shackle puts that kind of dynamic economics which introduces fixed time-lags into the reactions within a closed and determinate system free from uncertainty. Whether such a system has a tendency to equilibrium will depend on the exact numerical values ascribed to its parameters—small but critical changes here will transform negative into positive feed-back. Next

¹ Discussed in this *Journal*, 1955, 6, 66-78.

REVIEWS

comes the aggregative economics of the Keynesian type where uncertain expectations are important. Finally we have the economics of uncertain expectations itself.

The essays are written with verve, lucidity, and good temper. The author has a flair for metaphors with a (to borrow from his own terminology) high attention-arresting quality. Each essay takes the reader at once to the main problem and then swiftly and incisively develops the author's thesis.

J. W. N. WATKINS

Literature and Science.

Proceedings of the Sixth Triennial Congress of the International Federation for Modern Languages and Literatures. Oxford, 1954. Basil Blackwell, Oxford, 1955. Pp. 330. 45s.

THERE are fifty-seven papers in this fascinating volume, mostly in English or French. Among the topics dealt with in the papers are (a) the bearing of particular scientific studies, such as palaeography, bibliography, statistics, psychology, on problems of literature; (b) the influence on particular writers (e.g. Goethe, Balzac, Tchekhov, Proust, Valéry) of some particular science; (c) what kind and degree of awareness writers at a particular period had of the scientific doctrines of their time. There is an interesting account by Wladyslaw Folkierski of the borrowings by Fontenelle of passages in Galileo, for use in his *Pluralité des Mondes*, and of the relations between Fontenelle's work and the *Micromégas* of Voltaire. A. C. Crombie discusses Galileo's conception of Scientific Truth; Professor Dingle and Ronald Peacock show, each in his own way, that science cannot oust poetry; as Peacock puts it, there are 'two symbol systems that illuminate nature in complementary ways . . . either art, i.e. an interpreting image or science, i.e. a formula'. To add to the variety, there is an account by Gustave Charlier of the enthusiastic attitude of Belgian romantic poets to railway trains in the 1840's, and a paper by Miss Koutaissoff on the way in which Soviet writers of today think of themselves as instruments of propaganda for scientific knowledge and for the widespread use of technical inventions.

L. J. RUSSELL

Relativity for the Layman. By James A. Coleman.

The William-Frederick Press, New York, 1954. Pp. 131. \$2.75

THOUGH this *Journal* cannot concern itself with popular works, some readers may wish to know of a good popular book on relativity that they can safely recommend. Mr Coleman will give them what they want.

J. O. WISDOM

RECENT PUBLICATIONS ON THE PHILOSOPHY OF SCIENCE

(a) BOOKS RECEIVED FOR REVIEW

- Max Born, *Experiment and Theory in Physics*, Dover Publications, New York, 1956, pp. 43, 60c
- Max Born, *Physics in my Generation*, Pergamon Press, London, 1956, pp. viii + 232, 40s.
- Louis de Broglie, *Une tentative d'interprétation causale et non linéaire de la mécanique ondulatoire*, Gauthier-Villars, Paris, 1956, pp. vii + 297, 3,500 fr.
- Egon Brunswik, *Perception and the Representative Design of Psychological Experiments*, Cambridge University Press, London, 1956, pp. xii + 154, 37s. 6d.
- John C. Burnham, *Lester Frank Ward in American Thought*, Public Affairs Press, Washington, 1956, pp. 31, \$1
- Colin Cherry (Ed.), *Information Theory*, Butterworths Scientific Publications, London, 1956, pp. xii + 401, 70s.
- Orrin E. Klapp, *Ritual and Cult*, Annals of American Sociology, Public Affairs Press, Washington, 1956, pp. vi + 40, \$1
- Frederick Mayer and Frank E. Brower, *Education for Maturity*, Public Affairs Press, Washington, 1956, pp. vi + 155, \$3.25
- Frederick Mayer and Frank E. Brower, *Patterns of a New Philosophy*, Public Affairs Press, Washington, 1956, pp. vi + 112, \$3.25
- Maurice Natanson, *The Social Dynamics of George H. Mead*, Public Affairs Press, Washington, 1956, pp. vii + 102, \$2.50
- Bertrand A. W. Russell, *An Essay on the Foundations of Geometry*, Dover Publications, New York, 1956, pp. 201, \$1.50
- Gerold Stahl, *Introducción a la lógica simbólica*, Ediciones de la universidad de Chile, 1956, pp. xv. + 226
- Morris W. Travers, *A Life of Sir William Ramsey, K.C.B., F.R.S.*, Edward Arnold, London, 1956, pp. viii + 308, 50s.
- Curtis Wilson, *William Heytesbury*, The University of Wisconsin Press, 1956, pp. xii + 219, \$4.00
- Werner Wolff (Chairman of MD International Symposia), *Psychiatry and Religion*, MD Publications, New York, 1956, pp. 62, \$3.00

(b) ARTICLES

- A. J. Ayer, 'What is a law of Nature?', *Revue internationale de philosophie*, 1956, 10, 144
- Tomas A. Brody, 'Formación y extensión de los conceptos científicos', *Seminario de problemas científicos y filosóficos*, 1956, 11, 1
- G. Burniston Brown, 'Have we Abandoned the Physical Theory of Nature?', *Science Progress*, 1956, 44, 619-34